Journal of
Experimental Psychology

VOL. 27, NO. 3 SEPTEMBER, 1940

THE ORIGIN OF COMBINATION TONES [1]

BY ERNEST GLEN WEVER, CHARLES W. BRAY,
AND MERLE LAWRENCE

Princeton University

Distortion in the auditory mechanism is shown by the appearance in perception of two types of phenomenon, overtones and combination tones. Overtones appear when the ear is stimulated by a single sinusoidal tone of sufficient intensity. Combination tones arise when two or more such tones act simultaneously.

The problem of the origin of these phenomena, and especially the parts of the ear that are concerned, has received considerable attention.[2] Helmholtz [3] in his mathematical theory regarded them both as products of non-linearity in the movements of drum and ossicles. Others, like Schaefer,[4] attributed them to processes of the inner ear. A further possibility is that both middle ear and inner ear are concerned. Békésy,[5] largely on a basis of his experiments on the effect

[1] This investigation was aided by grants from the Rockefeller Foundation and the National Research Council.

[2] For a brief review of the problem, see E. G. Wever and C. W. Bray, Distortion in the ear as shown by the electrical responses of the cochlea, *J. acoust. Soc. Amer.*, 1938, **9**, 227–233.

[3] H. L. F. Helmholtz, *On the sensations of tone*, Trans. by Ellis, 2nd Eng. ed., London: Longmans, Green and Co., 1885.

[4] K. L. Schaefer, Eine neue Erklärung der subjectiven Combinationstöne auf Grund der Helmholtz'schen Resonanzhypothese, *Pflüg. Arch. ges. Physiol.*, 1899, **78**, 505–526.

[5] G. v. Békésy, Über die nichtlinearen Verzerrungen des Ohres, *Ann. Phys., Lpz.*, 1934, **20**, 809–827; Physikalische Probleme der Hörphysiologie, *Elektr. Nachr. Techn.*, 1935, **12**, 71–83.

of pressure on the eardrum, took the position that overtones arise in the inner ear, but combination tones are formed both in the middle ear and the inner ear. He observed that combination tones were altered by pressure on the drum, and concluded that the middle ear had a share in their formation.

In a recent report,[6] we have dealt with the problem of the place of origin of overtones. Evidence was presented that these effects are essentially independent of the peripheral conductive mechanism, and have their origin at some point beyond the stapes. The question which remains, and which forms the subject of the present paper, is whether combination tones have this same locus.

Method and Procedure

The method was similar to that previously employed in the study of overtones. Guinea pigs were used as experimental animals, and the electrical responses of the cochlea were taken as a measure of peripheral processes. The middle ear, up to the stapes, was surgically eliminated, and the activity of the remaining structures studied in isolation. Stimulation was effected by the application of a mechanical vibrator to the head of the stapes. A comparison of the pattern of combination tones obtained by this method with that secured from the intact ear indicates the locus of distortion.

As in the previous investigation, the point of separation between middle ear and inner ear was the joint between incus and stapes. The portions eliminated included drum, malleus, incus, the ossicular ligaments, and the tensor tympani muscle, and in 3 of the animals the stapedius muscle also.

The head of the stapes was driven by a steel needle attached through an insulating rod to a crystal vibrator. The vibrator consisted of a number of Rochelle salt sections 0.05″ thick by ⅝″ wide by 3″ long, cemented together to form a crystal of the expanding type ⅝″ × ⅝″ × 3″. It was mounted on a heavy reaction block and electrically shielded. A large manipulator was used to hold this instrument and to give accurate placement of the stimulating point.

Stimuli were produced by two beat-frequency oscillators, each provided with filters and attenuators, and arranged so that connection could be made either with the crystal vibrator for mechanical stimulation or with a loudspeaker for aerial stimulation. From the loudspeaker the sounds were conducted through a tube to a cannula that was securely tied in the meatus of the animal's ear.

For experiments of this kind it is essential that distortion be negligible in the stimuli themselves. Tests of purity were made for both conditions of stimulation. To the insulating rod of the vibrator was attached a vibration microphone, which consisted of a small Rochelle salt crystal with the inertia type of mounting. Tests were made when the vibrator was in its working position, with the point in contact with the stapes. For the aerial stimulus, a dynamic microphone fitted with a probe tube was employed. A sound-tight enclosure over the diaphragm of the microphone terminated in a tube small enough that it could be inserted in a hole in the side of the cannula while

[6] E. G. Wever, C. W. Bray, and M. Lawrence, The locus of distortion in the ear, *J. acoust. Soc. Amer.*, 1940, 11, 427–433.

the cannula was in position in the meatus. The sounds were thus picked up at the place of entrance to the ear, and usual conditions of acoustical impedance were preserved.

Tests for both conditions of stimulation showed combination tones in the stimuli to be negligible at all intensities employed.

The technique of applying the vibrator point to the stapes was the same as previously described. As a check upon whether the stapes faithfully followed the movements of the driving point, the pressure of application was varied. A moderate pressure was found to give regular responses, and an increase of this pressure had no discernible effect upon the distortion pattern until an extreme pressure was reached that caused damage to the ear.

The electrical potentials of the cochlea were recorded in the usual manner, with a silver foil electrode on the round window. The potentials, after amplification, were analyzed with a General Radio Type 736-A wave analyzer.

Experiments were made on 14 ears. In two instances the same ear was used first with aerial stimulation, in the intact condition, and later with mechanical stimulation after elimination of the middle ear, so that altogether 8 ears were investigated with each form of stimulation. Since with several ears more than one series of measurements were made, there were in all 10 series with aerial stimulation and 14 series with mechanical stimulation. Of the 14 series with mechanical stimulation 4 were with the stapedius tendon cut.

In most of the series the stimulating tones were 1000~ and 2800~. In a few, systematic measurements were made with 1000~ and 2400~, and with 1000~ and 3600~. No difference was found in the general pattern of combination tones produced by these three pairs of primaries.

Results

The experiments regularly included measurements on 8 tones, in addition to the primaries themselves. These were the two combination tones of the first order $h - l$ and $h + l$, the four tones of the second order $h - 2l$, $h + 2l$, $2h - l$, $2h + l$, and two tones of the third order $2h - 2l$ and $2h + 2l$. Frequently other tones were added, up to the fifth order, and occasional measurements were made on tones up to the eleventh order.

Table I shows the early orders of combination tones, with their algebraic relations and frequencies for primary tones of 1000~ and 2800~. All of these tones were found regularly, except the third order tone $3l - h = 200$~, which in some animals was not detectable. Why this tone was sometimes missing is not clear; in one of the animals in which it could not be recorded the use of other primaries, 2000~ and 5600~, gave the corresponding tone $3l - h = 400$~.

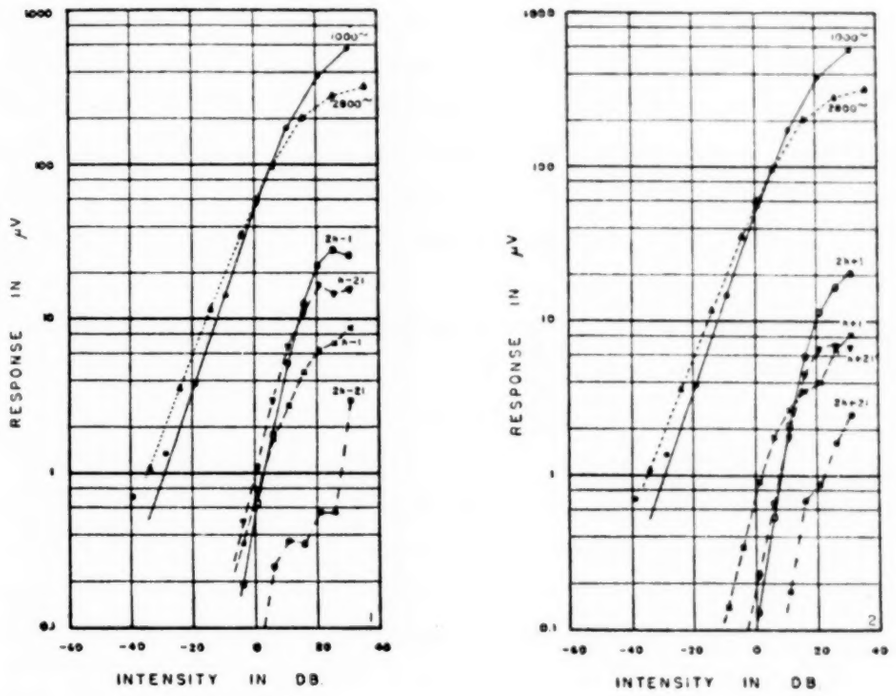
TABLE I

	h = 2800	$2h$ = 5600	$3h$ = 8400
l = 1000	$h - l$ = 1800	$2h - l$ = 4600	$3h - l$ = 7400
	$h + l$ = 3800	$2h + l$ = 6600	$3h + l$ = 9400
$2l$ = 2000	$h - 2l$ = 800	$2h - 2l$ = 3600	$3h - 2l$ = 6400
	$h + 2l$ = 4800	$2h + 2l$ = 7600	$3h + 2l$ = 10400
$3l$ = 3000	$3l - h$ = 200	$2h - 3l$ = 2600	$3h - 3l$ = 5400
	$3l + h$ = 5800	$2h + 3l$ = 8600	$3h + 3l$ = 11400

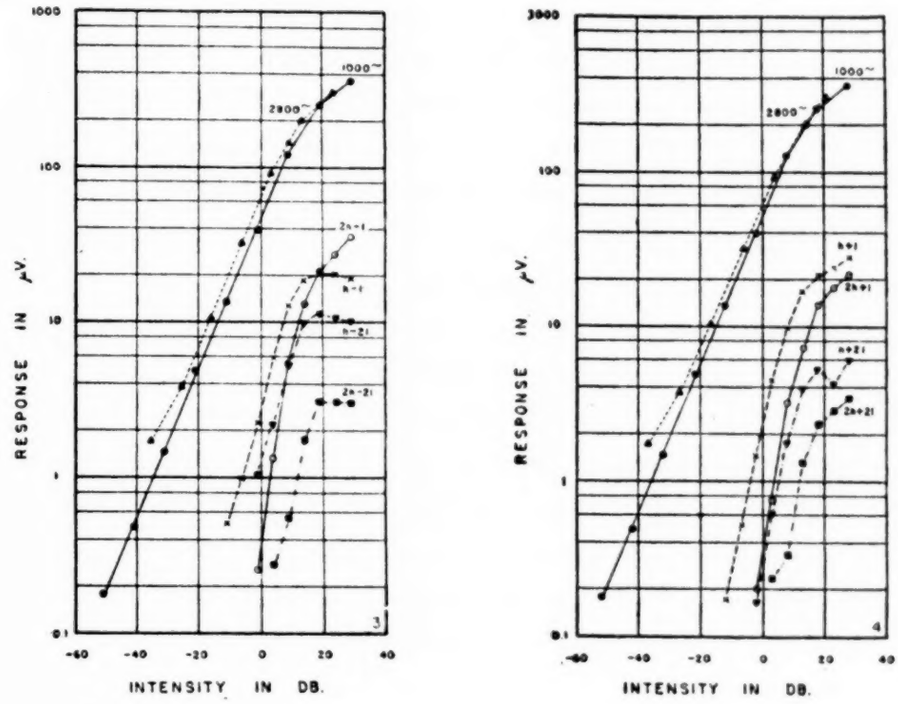
Figs. 1–10 show typical results for the 8 combination tones regularly studied. To reduce confusion, the difference and summation tones are given in separate graphs. In all the figures the ordinate shows the response in microvolts, and the abscissa the stimulus intensity in db. for the 1000~ primary. So far as the present investigation is concerned the reference intensity (zero db.) may be considered as arbitrary; actually it is a value of stimulus that gives the same response as ordinarily obtained with an aerial stimulus of 1 dyne per sq. cm. The 2800~ primary was adjusted to produce approximately the same response as the 1000~ tone, and then the two primaries were raised or lowered by equal steps (usually 5 db.) while the various combination tones were individually measured with the wave analyzer. On the graphs, the combination tones are indicated symbolically; the actual frequencies may be found in Table I. The curves for the primary tones show the magnitudes of response when these tones were presented singly.

Figures 1 and 2 show difference tones and summation tones respectively for aerial stimulation, with the ear intact. Figures 3 and 4 show these tones in another animal for mechanical stimulation after disarticulation of the stapes. If Fig. 1 is compared with Fig. 3, and Fig. 2 with Fig. 4, the general similarity of the results for the two forms of stimulation becomes apparent.

The combination tones first are measurable in the region of −10 to 0 db., and then they rise rapidly until they suffer



FIGS. 1 and 2. Difference and summation tones for aerial stimulation, with the ear intact. Guinea pig No. 293 R.



FIGS. 3 and 4. Difference and summation tones for mechanical stimulation of the stapes. Guinea pig No. 290 R.

bending at the high intensities for which the primaries themselves show bending. The combination tone functions all have slopes considerably greater than those of the primaries, which means that the combination tones increase in prominence as the stimulus intensity is raised.

The difference and summation tones show closely similar patterns, under all conditions. In many instances the difference tones exceed their corresponding summation tones, but only by small amounts.[7] The first and second order tones, as shown in these four figures, are of somewhat the same magnitude, while the third order tones, $2h - 2l$ and $2h + 2l$, appear at a higher intensity and attain a lower level of magnitude. In most of the animals studied, the first order combination tones have first been found at a slightly lower intensity than the others, but they tend to show a lower slope and relatively early bending, and usually are passed by the $2h \pm l$ components. This condition is shown clearly in Figs. 2 and 3, and to a limited degree in Fig. 1, but in Fig. 4 the $2h + l$ component falls below. Nearly always the $h \pm 2l$ component is exceeded by the $2h \pm l$ component; this is true in each of the four figures just presented.

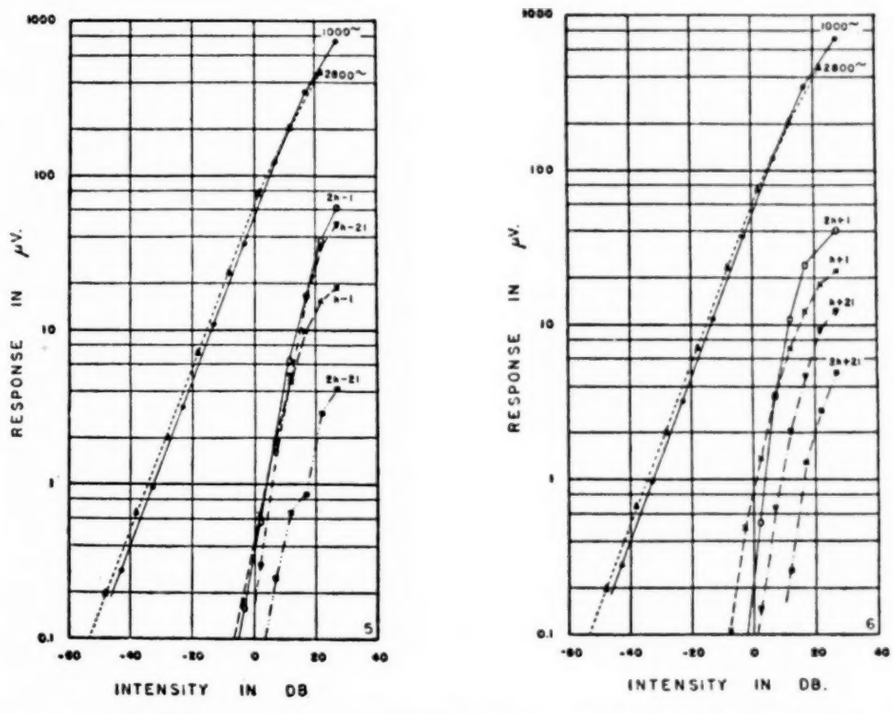
The combination tone curves often show irregularities, especially at the higher levels; this feature is possibly related to differences in the degree of overloading of the two primaries. It will be noted that though the primary curves run fairly close together at low and intermediate intensities, since they were adjusted for equivalence in that region, they often depart from one another at high intensities.

Most of the series, for aerial and mechanical stimulation alike, yielded combination tone patterns essentially similar to those shown in the preceding figures. Three of the series for aerial conduction, however, gave larger amounts of combination tones, particularly at the higher intensities of stimulation. The most striking of these series is here presented. It represents, also, one of the two animals in which measurements were made both with the ear normal and after disarticulation

[7] This relation of difference and summation tones was reported by E. B. Newman, S. S. Stevens, and H. Davis, Factors in the production of aural harmonics and combination tones, *J. acoust. Soc. Amer.*, 1937, **9**, 107–118.

of the incudo-stapedial joint. This complete procedure was not carried out as a rule, because the measurements require several hours, and it is difficult to maintain the animal in good condition for so long a time.

Figures 5 and 6 show difference and summation tones for aerial stimulation, with the ear normal. The responses are

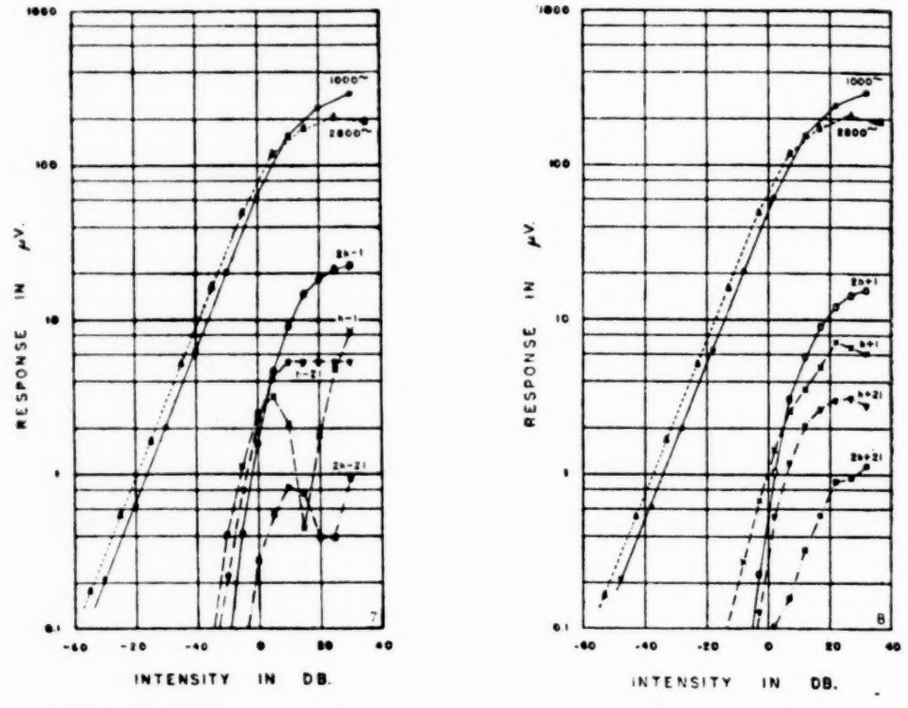


FIGS. 5 and 6. Difference and summation tones for aerial stimulation, with the ear intact. Guinea pig No. 294 R.

unusually strong, and the primary curves show relatively little bending at the highest intensities of stimulation. Likewise, the combination tone functions are fairly straight, and rise to comparatively high levels.

Figures 7 and 8 show results for the same ear after breaking the ossicular chain, with direct stimulation at the stapes. Some differences from the curves for aerial stimulation are apparent. The primary curves do not rise so high as before, and show considerable bending. The combination tone curves also bend over at high stimulus intensities, and in several instances become irregular. Perhaps the most significant difference between these and the preceding curves for aerial stimulation is brought out by measuring, at any given level

of response, the abscissa distance between the primary curves and any of the combination tone curves; the difference of 5 to 10 db. for the two forms of stimulation is a measure of the



FIGS. 7 and 8. Difference and summation tones for mechanical stimulation of the stapes. Guinea pig No. 294 R.

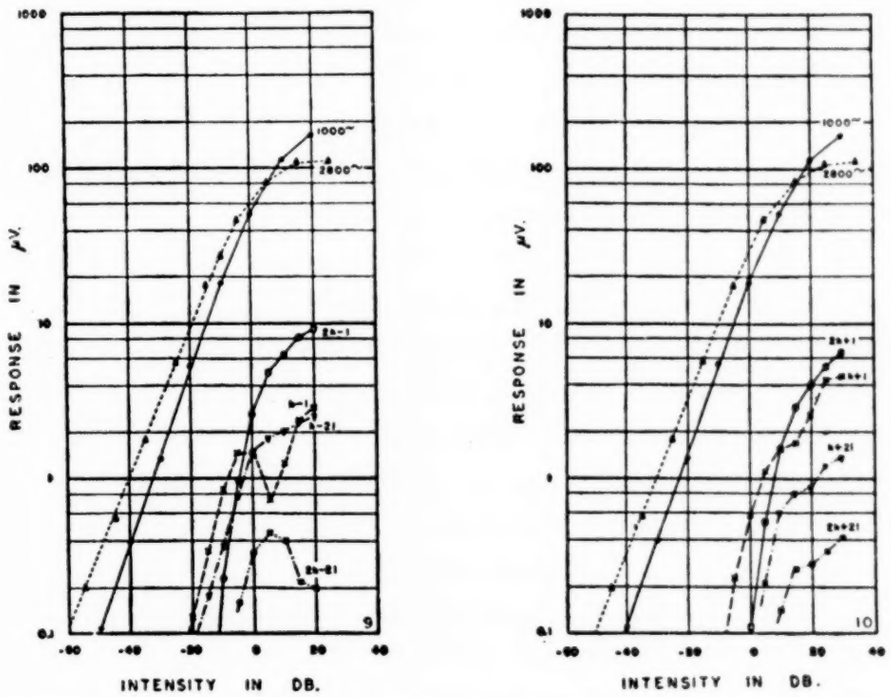
larger combination tone content for aerial stimulation in this animal.

The results just presented are somewhat difficult of interpretation. They may indicate that under some circumstances the middle ear makes a significant contribution to the total distortion pattern. Our hesitation in accepting this view comes from a consideration of the change in the form of the primary curves, and the likelihood that this change represents a change in the condition of the animal. A possible change is the greater accumulation of lymph about the niche of the round window. Nearly always, after the middle ear operation, this region is noticeably more moist than usual, due to the inevitable traumatizing of the tissues. It is usually found, under these conditions, that a careful drying of the region will raise the level of responses by 5 db. or so, for a short time. It is therefore possible that this animal was in unusually good

condition in the beginning, and suffered deterioration as time passed. Otherwise it would be difficult to explain why larger differences were found between the two forms of stimulation in this case than were ordinarily shown when different animals were used.

In any event, it is clear that the inner ear system alone, stimulated through the stapes, yields a distortion pattern that is generally similar to that of the ear as a whole.

In Figs. 7 and 8 the stapedius tendon was intact. It was then severed, and results were obtained which are shown in Figs. 9 and 10. As indicated, there was a further drop in the



FIGS. 9 and 10. Difference and summation tones for mechanical stimulation of the stapes, with the stapedius tendon severed. Guinea pig No. 294 R.

level of the responses, but the combination tone curves were altered but little. The evidence shows that the stapedius muscle is not significantly concerned in the production of the phenomena under consideration.

CONCLUSIONS

The results show, in general, a high degree of similarity in the combination tones produced by stimulation of the intact ear with aerial sounds and those produced by stimulation of the operated ear with a mechanical vibrator on the stapes.

The results do not rule out any participation by the middle ear in the combination tone pattern, since in some instances the effects were 5 to 10 db. greater when the ear was intact. As already brought out, technical conditions complicate the situation somewhat, but if these are set aside the results fall into agreement with Békésy's observations on the effect of pressure on the eardrum. As mentioned above, he concluded that combination tone distortion occurs both in the middle ear and in the inner ear.

The present results show clearly, in opposition to the view expressed by Helmholtz, that the middle ear is not the sole seat of combination tone distortion. Rather, they throw the principal emphasis upon the inner ear as the place of origin of these phenomena. They are thus in line with the position taken by Schaefer, Meyer, Bingham, and Peterson among earlier writers, and by Lewis and Reger, Békésy, Stowell and Deming, and Cotton in more recent times.[8]

The results are of course limited to the ear of the guinea pig under deep anesthesia. Our expectation is that, for the waking state, heightened tonus or active contraction of the middle ear muscles should reduce the amount of distortion rather than increase it. The evidence, previously reported,[9] is that tension on the tensor tympani tendon reduces transmission through the middle ear and diminishes distortion in strict relation to the change effected in the primary tones.

The particular process of the inner ear by which the combination tones arise is not further indicated by the evidence. A possibility is the process, attributed to the hair cells of the basilar membrane, by which mechanical energy is converted into electrical energy. This hypothesis has already been advanced in relation to overtone distortion, and it is reasonable to suppose that the same action is concerned in the production of combination tones.

(Manuscript received April 22, 1940)

[8] For the references, see J. Peterson, Combination tones and other related auditory phenomena, *Psychol. Monog.*, 1908, **9**, No. 39, pp. 133–136, and the article cited in note 2, above.

[9] E. G. Wever and C. W. Bray, The tensor tympani muscle and its relation to sound conduction, *Ann. Otol., etc., St. Louis*, 1937, **46**, 947–961.

THE ROLE OF MUSCULAR TENSION IN THE COMPARISON OF LIFTED WEIGHTS

BY BRYAN PAYNE AND R. C. DAVIS

Indiana University

From discussions of weight judgment by Needham (**7**) and Woodworth (**8**) one may infer that the theory proposed by Müller and Schumann (**6**) and its alternative, the physiological trace theory, have been most popularly employed in the interpretation of the problem of successive comparison. According to this theory, the subject adjusts the force of his lift to the first stimulus, carries over this adjustment to the second stimulus, and lifts it with a force about equal to that found necessary for the first. The resulting impression, then, is a function of the relation between inter-lift 'Einstellung' and the lift of the second stimulus.

In view of the kind of evidence, however, upon which this theory purports to rest, it seems proper to inquire whether, in its present state, it is more of a logical construct than a scientific concept based upon contacts with an observed datum. It appears that all successive comparison theory has been derived from psychophysical data, and in no literature appears what might be called *direct* evidence for the construction of such an hypothesis. Experiments performed in other sensory fields in which the time error has been found to increase with the lengthening of the *S–V* interval have contributed no support, for they have assumed, along with those of Müller and Schumann, that the fading trace or adjustment is there and that it is a function of the time interval. Somewhat better support of the hypothesis was found by Guilford and Park (**5**), who reported that the *PSE* was displaced in the predicted direction when the *S–V* period adjustment was changed with the interpolation of weights. Here, again, the evidence is not direct.

Now it appears rather certain that the term 'Einstellung'

was meant to indicate some sort of motor disposition of central nervous tissue [1] which governs the suspension of muscle excitation in the lifting arm or hand.[2] Furthermore, this 'Einstellung' is said to be adjustable.[3] If, however, it can be demonstrated that those properties which have been imputed to Einstellung are characteristic of muscular activity during the judgment of weight pairs, then it would seem that a strong case for a peripheral theory has been stated. It follows that a peripheral theory based upon the Müller-Schumann assembly of relationships would predict that when the degree of inter-lift muscular adjustment is high the judgment would be *L*, since the second stimulus requires relatively less effort to be lifted. By the same reasoning, when the degree of adjustment is low the predicted judgment would be *H*.

The development in recent years of refined techniques for the detection of minute muscular activities has provided a way of checking this theoretical structure against experimental fact. If the perception of lifted weights is a function of inter-lift set, and if set is a muscular phenomenon, then we should be able to demonstrate the differentials involved with the use of these delicate recording techniques. Davis (4) has recently reported an electromyographic study of the simple reaction time experiment in which he found that muscular phenomena appeared to demonstrate the properties imputed to set. This study has rendered the present approach to the problem an even more reasonable one.

Apparatus and Technique

One male subject, naive with regard to the weight lifting experiment, was given 50 preliminary trials with the use of the method of constant stimuli. The weights employed were 84, 88, 92, 96, 100, and 108 grams rotated upon an Urban Table. Each weight was increased 12.5 grams by a brass electrode and small wire lead, providing a signal system to be described in a later paragraph. The standard weight was 100 grams. The subject was trained in the following procedure: Two seconds after a 'ready' signal from E, O lifted the standard and replaced it within two seconds. At the end of

[1] *Cf.* (6, p. 44). "The nerve centers, in which the motor Einstellung happens. . . ."

[2] *Ibid.* (p. 37). "Bevor wir die Wirksamkeit dieser Einstellung an der Hand der Versuche näher darthun. . . ."

[3] *Ibid.* (p. 48). "The Einstellung can accordingly be adjusted to maintain a transmitted motor innervation and muscular activity. . . ."

the next three seconds, during which E rotated the variable weight into position, O lifted the variable and replaced it within two seconds. Two categories of judgment, *H* and *L*, were allowed. O was encouraged to report his judgment as quickly as possible, although the time required for this was allowed to vary. Approximately 15 seconds separated the successive presentations of weight pairs. A metronome beating once a second was present throughout the series. E's movements and the table of weights were obscured from O by means of a screen through which O projected his arm. The lift was largely a wrist movement, and during the time O's hand was not in contact with the weights it was suspended approximately one inch above them.

In the experimental series, O was presented with 173 weight pairings in random order. The pairings were presented over a period of six days, with not more than 33 pairings in any one day. Prior to each daily run, O was allowed five practice trials. During the recorded series, O sat in a shielded cage and was visually separated from the weights by means of a heavy screen. The active electrode was in surface contact with the dorsal side of the forearm (right) at the point of greatest enlargement, and through a conduit was connected with amplifying and recording systems in an adjoining room. The neutral electrode, placed upon the upper right arm, consisted of a thick felt pad dampened with saline solution. The amplifying and recording systems used in this study have been described in greater detail by Davis (**1**, **2**, **3**). The action potential generated in muscle tissue is picked up by the electrodes, fed through four stages of amplification, and impressed upon a five-inch cathode ray tube, upon which it may be seen as a small, bright blue spot in rapid oscillation. These oscillations are recorded upon photosensitive paper by means of a high speed moving film camera. Overall sensitivity of the recording system in this study was 3.66 microvolts input per .02 inch of film deflection. Amplification gain was linear down to 12 c.p.s., a figure well under that required in the present study.

Recorded in parallel with the action current were a time line in .01 second units and a signal line by means of which one was able to determine to within .01 second the points of lift and replacement of weights. To provide the signal system, each of the weights was fastened to a brass electrode weighing 12.5 grams and was then placed upon a small brass plate attached to the surface of the table. The weights and plates were wired in series with a signal key, used by E to indicate when the judgment was given, and with a second oscillograph in the recording room. When a weight was lifted, the break was recorded upon the film as a large deflection of the second oscillograph. This deflection was repeated when the weight was replaced. This signal technique has been described in greater detail by Davis (**4**).

A continuous action current record was taken of each of the 173 pairings in the experimental series. One second prior to E's 'ready' signal a second operator in the recording room signaled to E that recording had commenced. E then signaled to O, and the pairing was carried out according to the method described. Of the resulting 173 sections of film, 46 were rejected for measurement because of faulty procedure, recording, or development. Remaining were 127 satisfactory sections, of which 77 were *H* judgments and 50, *L* judgments. These sections were distributed by weight in the following way: 84 grams (7 *H*, 17 *L*), 88 grams (12 *H*, 11 *L*), 92 grams (19 *H*, 12 *L*), 96 grams (19 *H*, 5 *L*), 100 grams (15 *H*, 5 *L*), and 108 grams (5 *H*). Each section representing the record of one pairing was then marked off into the following periods:

FP—Foreperiod, consisting of three seconds prior to *PS*
PS[1]—One tenth second prior to standard lift
S—Standard lift

S–V—Period between standard lift and variable lift
PV[4]—One tenth second prior to variable lift
V—Variable lift
J—Period between weight replacement and judgment report
PJ—Period following report of judgment.

Each period was then marked off into .1 second units, and the greatest peak-trough difference in each .1 second unit was measured in fiftieths (.02) of an inch. The .1 second unit here was arbitrary, the only requirement being that the unit should not be so large that it would obscure differential trends in action currents. Over 14,000 measurements [5] expressed as microvolt input constitute the basis for the statistical operations of this study.

Experimental Results

In the main, two questions may be asked of the data obtained in this study. First, can differential characteristics of muscular action, such as would be demanded by a peripheral theory, be found when the data are separated upon the basis of *H* and *L* categories? Second, assuming that such characteristics are found, what is the predictive efficiency from them to the judgment category?

I. *Characteristic Muscular Action By Category*

A. Inter-categorical Differences.—In Fig. 1 is shown the entire course of microvolt input when all weight pairings are combined upon the basis of *H* and *L* categories. The stations shown here are means of comparable tenths of a second plotted against the temporal position of their occurrence in the judgment sequence. It may be noted first that when the judgment is *H* the course of input on the average lies *under* that for *L* judgments almost entirely up to the point of the second lift. Here it becomes very superior for the duration of the second lift, and then decreases in an inconsistent trend on down to the end of the sequence. With the exception of the lift periods and a few points in the *S–V* period the differences between these means are not statistically reliable. But we may think of any given tenth of a second as being subject to an enormous amount of chance variation that will tend to mask the reliability of any one difference. The reliability of

[4] The *PS* and *PV* periods are differentiated from the others because the movement here is unlike that of the lift periods, the foreperiod, or the *S–V* period.

[5] Nearly half these measurements were made by NYA assistants and have been thoroughly checked by the writer.

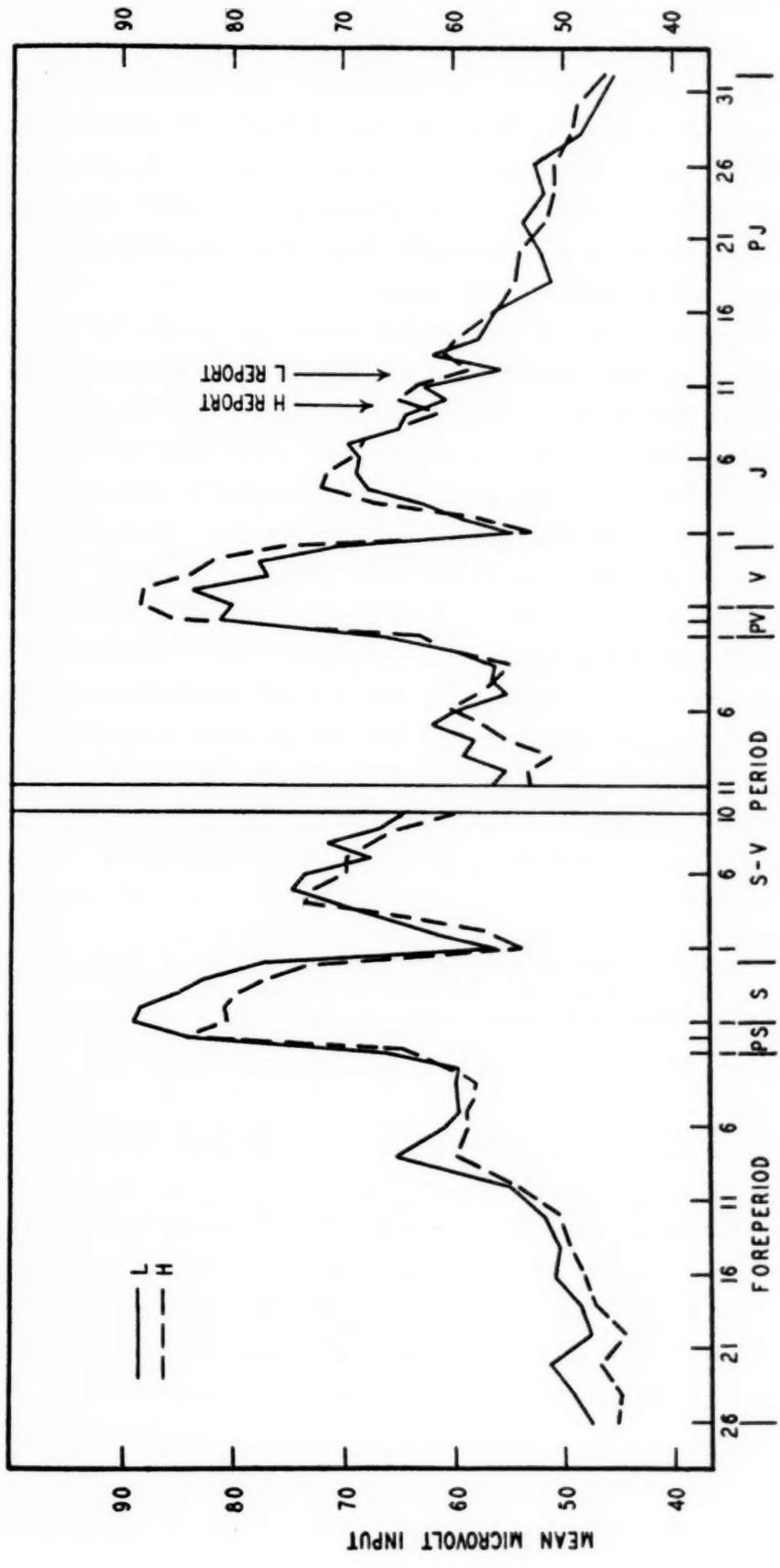


FIG. 1. The courses of microvolt input for *H* and *L* judgment sequences.

the differential trend is unmistakable, particularly in the foreperiod. If there we consider that at any one station the probability is 1/2 that the *H* mean will fall below the *L* mean by chance, we may ask: What is the probability that by chance it will continue to fall below for 14 consecutive stations? Expressed in this manner, the probability is $(1/2)^{14}$ or approximately six chances in 100,000 that the consistency of the entire trend is a chance function.

Perhaps it may be suggested that, in terms of the final interpretation, the per-tenth-second analysis is not a fair treatment of the data, since the significance of any one such unit is relatively small. The proposition that the effect of one period of muscular activity upon a later one is more of a summation effect is probably nearer the truth. Accordingly, in examination of the data from this point of view, a grouping technique was employed. The *FP*, *S–V*, *J*, and *PJ* were each divided into three approximately equal parts, and the measurements therein, categorically separated, were averaged. In addition, averages were found for the pooled first five .1 seconds of each lift period for each category, as well as for the *PS* and *PV*. Inter-categorical differences and reliabilities of differences were computed, as shown in Table I.

TABLE I

INTER-CATEGORICAL DIFFERENCES BETWEEN ACTION LEVELS WHEN ALL WEIGHTS ARE COMBINED

	FP_1	FP_2	FP_3	*PS*	*S*	SV_1	SV_2	SV_3	*PV*	*V*	PJ_3
Mean											
H	45.01	48.06	58.61	84.20	78.07	66.11	56.40	57.97	85.77	83.57	49.97
L	47.26	49.51	60.96	84.03	84.43	67.15	59.18	59.66	81.11	77.80	50.24
σ											
H	11.60	12.96	15.34		12.44	15.48	13.54	14.38	15.04	13.80	12.04
L	12.66	12.70	13.65		14.16	15.08	14.75	13.65	14.71	12.37	11.35
N											
H	750	759	760	77	385	646	652	642	77	385	541
L	498	502	499	50	250	409	414	413	50	250	279
Diff.	2.25	1.45	2.35	.17	6.36	1.04	2.78	1.69	4.66	5.77	.27
σ Diff.	.71	.74	.83		1.10	.96	.89	.88	2.69	1.05	.85
$\frac{\text{Diff.}}{\sigma \text{ Diff.}}$	3.17	1.96	2.83		5.78	1.08	3.12	1.92	1.73	5.50	.32

The clarity of the differences is increased by this grouping technique. The foreperiod action level for *L* judgments is

found to be generally reliably higher than that for *H* judgments, as is the central third of the *S–V* period. It will also be noted that when the judgment is *L* the tensional level of the standard lift period is reliably higher than that for *H* judgments, whereas in the variable lift period the tensional level for *L* judgments is reliably lower than that for *H* judgments. In the *PJ* period it will be noted that a marked difference occurs between the two trends at the 3rd, 4th, and 5th .1 second periods. If we combine all the *H* measurements at these points we find an average input of 72.19 microvolts as compared with the average of 68.17 microvolts for similarly combined *L* measurements. The ratio of this difference to its sigma is 2.63. Other than this, not much of significance can be noted in the two *J* and *PJ* trends, except that the tensional levels for both categories drop gradually down to the initial levels of the foreperiods. This demonstrates only that in these data there are no indications of muscular effects from one pairing to another.

The preceding evidence has demonstrated significantly different trends in microvolt input which are related consistently to particular categories. It may be proper to inquire now whether this holds true when the data are classified as to weight as well as to category. When the statistical scores derived from data grouped as in Table I are summarized by weight it is noted that the only reversal of the general findings previously indicated lies in the *S–V* period for 84 grams. This reversal, however, is not statistically reliable.

B. Intra-categorical Differences.—Ultimately, in describing what happens peripherally when weight pairs are judged as they are, one should perhaps more properly point out intra-categorical trends than inter-categorical differences. Presumably, the *heavier-than-ness* or *lighter-than-ness* of the variable stimulus is a function largely of those events which have preceded it in the particular sequence of which it is a part. If these events are muscular in nature, then one should be able to find them.

Statistical summaries of the data obtained from the two lift periods for both *H* and *L* categories are shown in Table II.

TABLE II

INTRA-CATEGORICAL DIFFERENCES BETWEEN STANDARD AND VARIABLE LIFT POTENTIALS BY WEIGHT AND TOTAL

	84 gm.		88 gm.		92 gm.		96 gm.		100 gm.		Total	
	H	*L*	*H*	*L*	*H*	*L*	*H*	*L*	*H*	*L*	*H*	*L*
Mean												
S	81.80	84.73	78.62	83.67	76.31	86.86	77.95	87.77	78.91	83.81	78.07	84.43
V	86.08	78.32	87.77	76.49	82.28	79.67	83.39	72.83	83.96	79.28	83.57	77.80
σ												
S	13.83	14.46	11.68	16.36	12.70	12.66	11.46	10.21	12.59	12.33	12.44	14.16
V	12.63	11.46	12.70	12.41	14.68	12.19	12.22	13.83	14.16	13.47	13.80	12.37
N												
S	35	85	60	55	95	60	95	25	75	25	385	250
V	35	85	60	55	95	60	95	25	75	25	385	250
Diff.	4.28	6.41	9.15	7.18	5.97	7.19	5.44	14.94	5.05	4.53	5.50	6.63
σ Diff.	3.17	2.00	2.22	2.76	1.99	2.26	1.71	3.43	1.28*	2.26*	.94	1.18
$\frac{\text{Diff.}}{\sigma \text{ Diff.}}$	1.35	3.21	4.12	2.60	3.00	3.18	3.18	4.36	3.95	2.00	5.81	5.58

* Coefficients of correlation were considered here.

The means shown here are derived from the pooled first five .1 seconds of a given lift (*S* or *V*) of a given weight of a given category (*H* or *L*) of judgment. In the last column (Total) the measurements are pooled according to lift and category alone. It will be observed that when the judgment is *H* the action level during the variable lift is usually reliably higher than that found during the standard lift. The converse occurs when the judgment is *L*, and it occurs without regard to the weight pair being compared. The only exception to this finding lies in the *H* judgments for the 84 gram pairing, and this exception involves only a reliability of a difference.

The computation of ratios between certain periods in the pairing sequence provides another method of examining the data for intra-categorical trends. Upon the basis of comparisons of absolute tensional levels in the lift periods, as in Table II, we are led to suppose that certain relations will be found with the use of these ratios. For example, if the mean action level of the variable lift period is lower than that of the standard lift period for *L* judgments, we should find that the mean ratio of *S* to *V* (*S* divided by *V* for each pairing) is larger than that for *H* judgments. That this is the case is amply

shown in Table III. The mean of the S/V ratios for L judgments is reliably larger than that for H judgments by a critical ratio of 8.02.

TABLE III

INTER-CATEGORICAL DIFFERENCES BETWEEN MEANS OF CERTAIN INTRA-CATEGORICAL RATIOS

	Ratio S/V	Ratio FP/S	Ratio SV/S	Ratio SV/V
Mean				
H	94.70	56.86	71.64	67.30
L	109.30	56.44	69.92	75.90
σ				
H	9.35	10.41	12.99	13.10
L	10.51	11.16	14.03	15.15
N				
H	77	74	77	77
L	50	50	50	50
Diff.	14.60	.42	1.72	8.60
σ Diff.	1.82	1.98	2.48	2.61
$\frac{\text{Diff.}}{\sigma \text{ Diff.}}$	8.02	.21	.69	3.29

Now it appears that the standard lift period potentials are categorically different. To explain this difference we are led to suppose that these two levels are functions largely of their preceding foreperiod levels. It follows, then, that FP_L should bear the same relation to S_L as FP_H does to S_H, and that there should not be a reliable inter-categorical difference between the mean ratios of foreperiod to standard (FP/S). Accordingly, the FP/S ratios, computed for each pairing, were categorically separated and averaged. Table III shows the ratio of this difference to its sigma to be only .21.

The situation is similar with regard to the relation of S–V to S. If the two S–V levels are functions of their respective standard periods, then S–V_L should bear the same relation to S_L as S–V_H does to S_H. When the ratios of S–V to S are computed for each pairing, categorically separated, and averaged, it is found that the difference between means is not reliable. The critical ratio of .69 shows that on the average the S–V for L judgments recedes from its standard period in the same proportion that the S–V for H judgments recedes from its standard period.

Finally, since it was found that on the average a rela-

tively lower *V* level followed a relatively higher *S–V* level and conversely, it should appear that the mean ratio of *S–V* to *V* for *L* judgments is reliably higher than that for *H* judgments. Table III shows the critical ratio here to be 3.29. Apparently it has been demonstrated with the use of ratios that the *S* and *S–V* periods for *H* and *L* judgments are proportional to their immediately preceding periods, and that the *V* periods, on the average, bear an inverse relation to their respective *S–V* periods.

II. *Prediction from Action Current to Judgment*

In an attempt to answer the second question one may control statistically the level of muscle action and observe within these controls the predictive efficiency from action level to judgment. It was noted in Fig. 1 that the action curve for *H* judgments undercut that for *L* judgments generally up to the *V* lift periods. Therefore, if we should combine into a single distribution all the *H* and *L* measurements obtained from a given period in the pairing sequence we should find that a greater proportion of *L* judgments is predicted from the upper extreme of the distribution than from the lower extreme. It follows that if an *L* judgment is predictable from a high action level the two extremes of any of the distributions should differ reliably in the proportions of *L* judgments which they predict. For the examination of this proposition we computed the percentages of *L* judgments predicted from the extremes of the *FP*, *S*, and *S–V* distributions and of the *S/V* ratio distribution. These are shown in Table IV together

TABLE IV

PREDICTIONS OF L JUDGMENTS FROM EXTREMES OF COMBINED H–L DISTRIBUTIONS AND FROM *S/V* RATIO DISTRIBUTION

	Foreperiod		Standard		*S–V*		*S/V* Ratio	
	N	% *L*	*N*	% *L*	*N*	% *L*	*N*	% *L*
Upper	409	42.00	131	58.00	522	44.00	32	81.00
Lower	405	31.00	132	28.00	556	34.00	34	6.00
Diff.		11.00		30.00		10.00		75.00
σ Diff.$_p$		3.35		5.81		2.95		8.00
$\frac{\text{Diff.}}{\sigma \text{ Diff.}_p}$		3.28		5.16		3.39		9.37

with reliabilities of differences between proportions. It appears from these results that reliably differential prediction may be made from the extremes of these distributions. The general principle indicated here holds true whether we combine measurements from an entire period in the sequence or deal only with measurements at a given tenth of a second.

III. *The Transition Point and V/S Ratios*

The analysis of the data thus far has indicated that the perception of lifted weights is most likely a function of the relation between action potentials during standard and variable lifts. This has been demonstrated regardless of whether we proceed from category to action level (Table II) or from action level to category (Table IV). Now if we consider further the ratio of each variable lift to its standard lift, a more critical examination of this proposition becomes possible. A V/S ratio of 1.00 would seem to indicate that, as far as concomitant muscular action is concerned, there is no heavier-than-ness or lighter-than-ness. In other words, it would indicate a transition point in muscular action. If H and L judgments are functions of V/S ratios, then the transition point in muscular action should correspond closely to the transition point in judgment, the 50 percent point.

In order to test this proposition, we computed the mean V/S ratio for each weight employed [6] and plotted it against its respective weight, as in Fig. 2. Upon the same set of axes was plotted the psychometric curve of H proportions. Now if we drop perpendicularly from the ratio curve at the point $V/S = 1.00$ to the abscissa we find that the point of intersection is equivalent to 86.25 grams. By interpolation of the median upon the p-curve we find a transition point of 87.65 grams. These two computations, of course, assume linearity between the points involved on both the ratio and p-curves, and it may well be that the assumed linearity is responsible for the fact that the computed weights do not exactly coincide.

Interesting enough in its own right is the form of the ratio curve and its relation to that of the p-curve. In general it

[6] Since there were only five measurements for it, the 108 gram weight was excluded.

appears that the ratio curve rises as the *p*-curve, but with greater negative acceleration. This 'diminishing returns' feature suggests as one possible interpretation that we are dealing here with what may be the upper portion of an ogive, and that

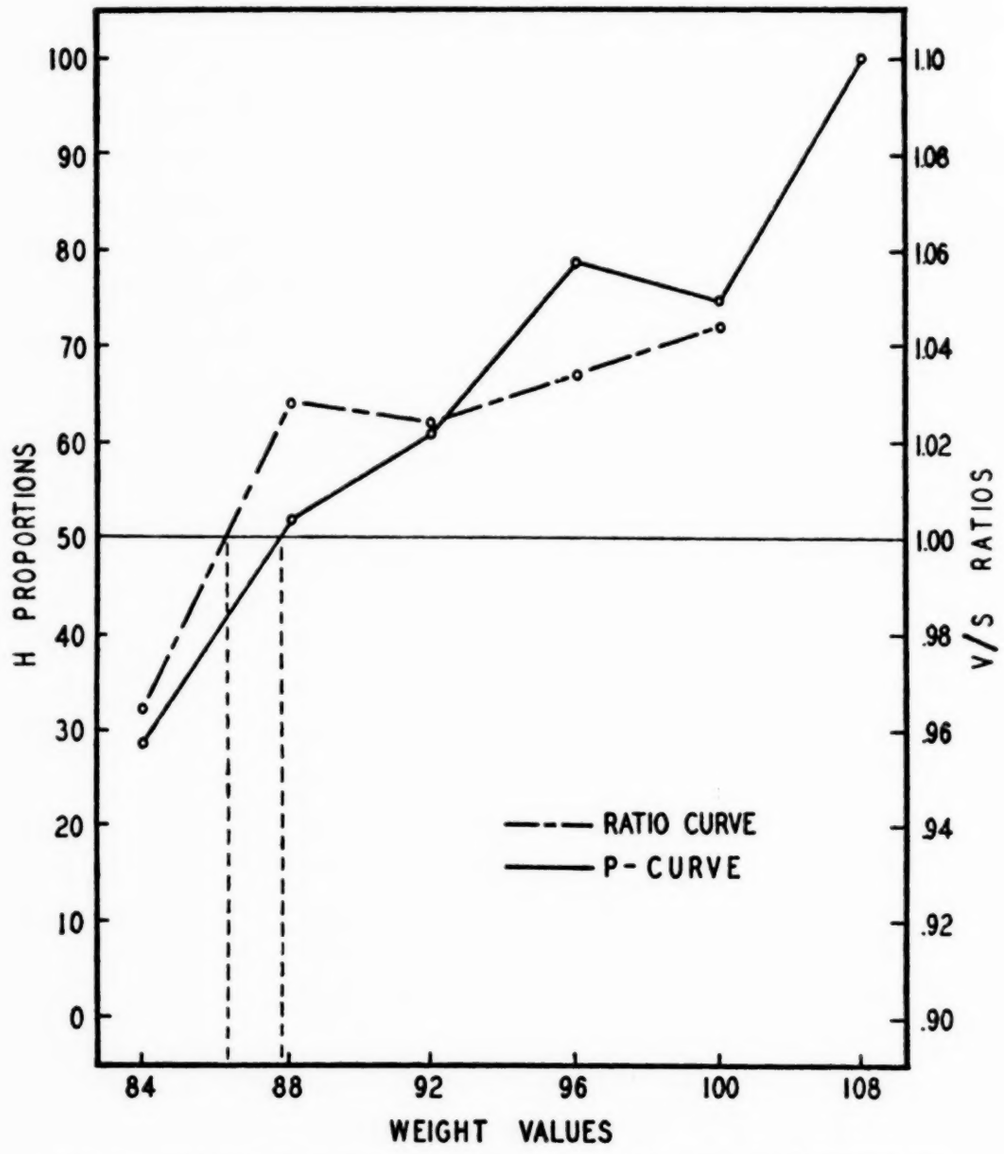


FIG. 2. The *V*/*S* ratio curve and the psychometric curve of *H* proportions.

the hypothetical lower portion was obscured because the weight stimuli did not extend far enough below this particular transition point. If this should be the case, it might mean that the difference in muscular action (between standard lift and variable lift) that is necessary in order to be perceived as different follows a law of diminishing returns on either side of

the transition point. This may represent a lag in muscular activity as compared with the stimulus weight, or—possibly—the sensation as progressing more slowly than the stimulus. At any rate, the data are more suggestive than conclusive.

Discussion of Results

It appears that the results of this study have shown that the properties imputed to Einstellung (central disposition) by Müller and Schumann serve to characterize muscular activity during the perception of lifted weights. To the extent that they do characterize this activity they would seem to indicate a peripheral theory based upon the assembly of relationships to be found in the Müller-Schumann proposal. In his discussion of peripheral theory, Davis (**2**, page 5 f.) suggests that "it might require the occurrence of different patterns of muscular activity for different psychological processes, and, conversely, of the same motor pattern for the same psychological processes on all occasions." Clearly these results have shown, not simply a general muscular accompaniment of this judgment process, but differential trends in activity that are related consistently to particular categories. These differential trends were found, not only when the analysis proceeded from category to action level, but also when it proceeded from action level to category.

Further examination of the relationships between the various periods within the pairing sequence revealed that the greatest predictive efficiency from action level to category was yielded by the extremes of the V/S ratio distribution. The hypothesis was then suggested that the transition point in the ratio series (V/S = 1.00), when interpolated into the weight series, should approximate the transition point interpolated upon the p-curve for H judgments. The suggested close approximation was found. The major fact, therefore, which seems to require interpretation is that, as far as muscular phenomena are concerned, the judgment is a function of the ratio of variable lift to standard lift.

Now it is difficult to see how the muscular action during the variable lift can be accounted for without reference to the

action that has preceded it in the pairing sequence. Therefore, let us inquire about the conditions obtaining at the time of the standard lift. It appears here that the lift of the same stimulus weight (100 grams) produces tensional levels which, when separated according to their correlated categories and averaged, are found to be reliably different. This difference would be an exceedingly difficult one to rationalize were it not for the supporting fact that the higher standard lift is preceded by the higher foreperiod level. The interpretation is now suggested that the *L* standard lift is higher than the *H* lift because the stimulus weight intrudes upon a higher level of activity already in progress. That this assumption is not wholly without support is indicated by Davis (1, page 20 f.) who, working in an entirely different setting, concluded, "Previous activity favors larger responses, previous quiescence small . . . the muscle tension response . . . is not subject to direct voluntary control, but is facilitated or inhibited, or both, by previous conditions." If this is the case, then the conditions-leading-to-judgment are partially established as far back as the foreperiod.

Similarly, this same principle seems to account very well for the two action levels during the *S–V* period. It was shown in Fig. 1 and Table III that $S–V_H$ recedes in proportion to S_H as $S–V_L$ does to S_L; and nowhere in this period is there reason to suppose that the action trends are representing anything more than their respective preceding activities.

At the beginning of the variable lift, however, the state of affairs appears to be more complex and not readily resolved with the use of the aforementioned principle. As with the standard stimulus weight, the lifting of the same variable weights produces tensional levels which, when categorically separated and averaged, are found to be reliably different. But the curious fact is that conditions preceding the second lift are the reverse of those preceding the first lift, viz., that the lower lift is now preceded by the *higher* action level. One possible solution of this is that, perceived against a relatively lower *S–V* level, *V* will seem relatively heavier and to require more effort to be lifted. According to this, the *V* level would

be a partial function of some sort of perceptual activity engaged in by *O* at the beginning of *V*. In fact, this is necessary if the judgment is to bear any relation to the actual weight of the variable. It is clear that the Aufgabe forces the beginning of the *V* lift into prominence as the psychologically critical period of the sequence. We may say then that the magnitude of muscular action during the *V* lift is determined by two factors, viz., the preceding organismic condition (*S–V* level), and the size of the weight itself.[7]

Summary and Conclusions

The suggestion was made that popularly employed theories of the problem of successive comparison were derived, not from contact directly with an experimental datum, but indirectly from psychophysical data. In order for us to examine the possibility of a peripheral theory of the perception of lifted weights, we obtained a continuous electromyographic record of the performance of one trained subject during the judgment of 127 pairs of weights. Two categories of judgment, *H* and *L*, were allowed. Under the conditions of this study, the results seem to support the following conclusions:

1. Differential trends in muscular action were found when the data were separated as to weight-category and category.

2. Reliably differential predictability from action level to category was found.

3. Statistically reliable differences were found between the two lift periods within a judgment category, and between the two categories at comparable lift periods.

4. The conditions-leading-to-judgment are partially established as far back as the foreperiod.

5. There is no evidence of muscular effects from one pairing to another.

6. As far as muscular action is concerned, the judgment of lifted weights is a function of the ratio of variable lift to standard lift.

7. The absolute magnitude of muscular action during the

[7] See Fig. 2. The ratio of variable lift to standard lift increases with the size of the weight encountered.

standard lift is determined by two factors, viz., the preceding foreperiod activity, and the weight encountered.

8. The magnitude of muscular action during the variable lift is determined by two factors, viz., the organismic condition (S–V level), and the size of the weight itself.

9. The weight value which indicates equality in standard-variable muscular action corresponds closely to that which indicates equality in judgment.

10. The close correspondence between properties of muscular action during weight lifting and properties imputed to Einstellung and physiological trace suggests that these concepts may be reformulated into a peripheral theory of the perception of lifted weights.

(Manuscript received April 3, 1940)

References

1. Davis, R. C., The muscular tension reflex and two of its modifying conditions, *Ind. Univ. Publ. Sc. Ser.*, 1935, **3**, 21 pp.
2. Davis, R. C., The relation of certain muscle action potentials to 'mental work,' *Ind. Univ. Publ. Sc. Ser.*, 1937, **5**, 29 pp.
3. Davis, R. C., The relation of muscle action potentials to difficulty and frustration, *J. exp. Psychol.*, 1938, **23**, 141–158.
4. Davis, R. C., Set and muscle tension (In Press).
5. Guilford, J. P., and Park, D. G., The effect of interpolated weights upon comparative judgments, *Amer. J. Psychol.*, 1931, **43**, 589–599.
6. Müller, G. E., and Schumann, F., Ueber die psychologischen Grundlagen der Vergleichung gehobener Gewichte, (Pflüger's) *Arch. f. d. ges. Physiol.*, 1889, **45**, 37–112.
7. Needham, J. G., The time-error in comparison judgments, *Psychol. Bull.*, 1934, **31**, 229–243.
8. Woodworth, R. S., *Experimental Psychology*, New York, 1938, Henry Holt & Co., 392–450.

SOME FACTORS IN ESTIMATING SHORT TIME INTERVALS

BY A. R. GILLILAND
Northwestern University

(WITH THE COÖPERATION OF MR. RICHARD MARTIN,
Western Reserve University)

In an earlier study [1] attempts were made to discover some physiological process or processes which operate in estimating time. In this study pulse rate, breathing rate, heart work, lung work, and blood pressure changes were found to have no significant correlations with time estimations. In a preliminary study,[2] which has just been completed, no relationship was discovered between psycho-galvanic changes and time estimations. Gardner [3] found no relationship between basal metabolic rate and time estimations. The evidence presented by Francois [4] and Hoaglund [5] does not seem to justify the assumption of some king of internal time clock for the estimation of time. Although there may be some other physiological changes which form cues for time estimation, the evidence now available suggests that the factors in time estimation are mostly external, that is, time is inferred [6] from events taking

[1] Gilliland, A. R., and Schaefer, V. G. The relation of time estimation to certain physiological changes. *J. exp. Psychol.*, 1938, **23**, 545–552.

[2] The evidence in this preliminary study was not conclusive and this problem is receiving further study.

[3] Gardner, W. A. Influence of the thyroid gland on the consciousness of time. *Amer. J. Psychol.*, 1935, **47**, 698–701.

[4] Francois, M. Contribution à l'étude du sens du temp. La témperature interne; comme fasteur de variation de l'appréciation subjective des durées. *L'Année Psychol.*, 1927, **28**, 186–204; and Influence de la témperature interne sur *notre* appréciation du temps. *C.R. de la Soc. de Biol.*, 1928, **98**, 201–203.

[5] Hoaglund, H. The physiological control of judgments of duration: Evidence for a chemical clock. *J. gen. Psychol.*, 1933, **9**, 267–287; and The physiological control of judgments of duration. *Am. J. Psychol.*, 1934, **109**, 54.

[6] It is not to be understood from this statement that there are no incipient physiological correlates of time perception. There are certainly such in the sense that there are physiological correlates or equivalents of all mental processes.

place outside the body rather than from physiological processes of the type mentioned above going on within.

The principal study to be reported in this article is a comparison of the effectiveness of estimating time by counting with that of estimating time without counting. A study was also made of the possible effectiveness of estimating time by counting as compared with that of estimating time without counting after practice.

College students, when asked to estimate the length of short time intervals, almost invariably use some method of counting. In one group, which was typical, most of the subjects (82 percent) tried to count to themselves at the rate of one per second. Others (8 percent) tried to estimate the rate of movement of the second hand of a clock and count the seconds in this way. A few others tried to tap with their foot or nod their head once per second. A still smaller number tried to estimate the intervals by counting their heart-beats or breathing rate.

Is some kind of counting important or can time be estimated equally well without counting? This is the principal problem of the present study. Of course, it is at once apparent that counting is not essential in time estimation because we can make some kind of estimate of time that has passed when we were not expecting to make an estimate. For example, if we are asked how long since a bell has rung, we can often make a fair guess even though we have not been intending to estimate the time and certainly did not count. But can we estimate as well when we do not count, and, if not, how important is the counting?

Four different groups of college students were used as subjects in the study. There was a total of 301 subjects in the four groups: 70 in Group I, 32 in Group II, 100 in Group III, and 98 in Group IV. Groups I, III, and IV were from beginning classes in psychology and Group II from two advanced classes. The intervals which were to be estimated ranged from 4 to 27 seconds in length. With the first two groups ten intervals were used as follows: 4, 6, 8, 11, 12, 14, 18, 19, 22, and 27 seconds. With the third and last groups eight intervals were used: 4, 6, 8, 11, 14, 18, 22, and 27 seconds. These intervals were always given in mixed order. The time estimated was an unfilled interval bounded by two clicks of a telegraph sounder. The time between the clicks was measured by a silent seconds pendulum which was wired in circuit with the sounder and a switch key. All of this apparatus was hidden from the subjects by a screen.

The subjects were provided with a mimeographed sheet for recording their estimations. They estimated the length of each interval immediately after the last click bounding the interval.

In the first part of the experiment the students were told that they were to estimate the length of the intervals to the best of their ability. No mention was made concerning the method they were to use except that they were not to look at a watch or at their neighbor's report. After first series of intervals were presented, the subjects were asked to indicate what method they had used. Some of the common methods were described to the students, but it was pointed out that there were other methods than those mentioned. Since there was such a universal tendency for the subjects to count when no directions as to method were given, these trials will henceforth be called the counting series. Immediately following the first experiment the intervals were again presented but in another order. In the second estimation the subjects were directed not to count. Emphasis was placed upon not counting or using any method of filling in the time interval. The first group was asked whether they had succeeded in not counting and almost all of them believed they had not counted. The second group tested were advanced psychology majors. They were tested in two separate sub-groups, and close observation as well as their own reports indicated that except in a very few cases they did not count. The third and fourth groups were first allowed to count and then retested. Following the retest the subjects in the third group were asked to indicate on a five point scale whether or not they thought they had counted. The results indicated that they believed in almost every case that they did not count or use any internal factor of which they were aware in their estimating.

Of course this does not prove that the subjects did not count. Counting may have been unconscious and yet effective as an aid to estimating time. Since it is impossible for the mind to remain a blank while we are awake, some sort of material must have filled the interval between the two clicks of the sounder. It is also true that although this time was not filled by deliberate counting, it, no doubt, was filled with material which formed a basis for some kind of grouping. This grouping, whether conscious or unconscious, could well form the basis for the estimation of the length of the interval. While in a sense this might be considered only a poorer form of counting, it is so different from the deliberate counting that we seem to be justified in considering the two methods of estimating as essentially different—one in which there was counting, and the other so different in nature that we may well call it not counting.

About two weeks after the first test the experiment was repeated with all four groups. Altogether there were about 5,000 estimations made with counting and the same number without counting. Errors in estimation were determined by taking the difference between the correct time and the estimated time of each interval for each subject. The direction of each error was indicated by a plus sign for errors in which the estimated time was longer than the correct time and a minus sign when it was shorter. However, in the data presented here the direction of the error was not considered except in Table V. In all other cases, minus and plus errors were not differentiated. The percent of error was determined by dividing the amount of the error by the correct time. Other methods of indicating errors were considered and even tried out but were discarded in favor of this method.

The following tables (I–IV) give a summary of the results for each of the four groups used in the study. For all four groups the median percent of errors for the two tests with

TABLE I

70 CASES

Median Percents of Errors in Estimation

Time Intervals in Seconds	1 Counting		2 Not Counting		3 Counting		4 Not Counting	
	Error	Sigma	Error	Sigma	Error	Sigma	Error	Sigma
4	28.3	4.16	28.5	2.85	28.3	2.65	29.5	3.00
6	26.5	2.60	30.6	3.95	27.7	1.47	30.7	4.40
8	29.4	4.50	29.7	4.50	19.2	3.10	25.5	2.95
11	35.0	3.70	28.8	2.86	28.0	3.30	35.4	5.70
12	27.2	4.25	28.3	2.50	28.0	3.00	28.0	2.85
14	25.0	3.70	25.3	3.75	25.3	1.86	26.1	3.00
18	26.5	3.40	26.0	2.05	30.0	2.20	24.0	2.70
19	29.0	4.15	23.4	2.70	23.2	2.40	27.5	2.20
22	26.0	4.50	26.5	3.30	27.0	2.60	32.3	3.00
27	29.0	2.96	25.0	2.50	20.0	1.86	29.0	3.20
Combined median	27.2	1.18	27.6	1.00	26.7	1.00	28.2	.83
	C.R. = 0				C.R. = 1.15			

TABLE II

32 CASES

Median Percents of Errors in Estimation

Time Intervals in Seconds	1 Counting		2 Not Counting		3 Counting		4 Not Counting	
	Error	Sigma	Error	Sigma	Error	Sigma	Error	Sigma
4	29.4	2.85	27.1	4.80	27.5	4.80	27.7	3.90
6	32.5	4.40	30.0	4.25	27.1	3.10	27.5	3.50
8	30.0	4.80	28.0	5.60	25.0	6.25	26.5	7.60
11	28.3	4.50	37.0	5.55	26.0	4.75	29.4	5.80
12	27.1	6.10			25.0	5.55	24.1	8.90
14	25.0	6.70	29.0	6.30	26.7	4.90	30.8	8.00
18	27.5	5.00	26.7	4.45	21.0	4.75	37.5	7.40
22	25.0	6.50	30.0	6.50	24.0	3.50	30.6	8.00
27	29.0	5.20	31.0	4.60	25.0	7.80	33.7	4.70
Combined median	28.9	1.75	29.1	1.80	24.9	1.38	28.3	1.73
	C.R. = .08				C.R. = 1.54			

counting and the two without counting are given. Medians rather than means were calculated because, although the distributions were almost normal in the middle ranges, there were often a few large overestimations which would have had undue influence on the means. As already stated, in the first four tables the median percents of errors are calculated with-

out regard for the sign; that is, the errors in either direction are treated only as that amount of error. The data are arranged in order of length of the time intervals estimated, but, as already stated, they were not presented in that order for estimations. In these tables the median for each time interval is given separately, but the medians are also combined into a general median for each counting and not counting series.

TABLE III

100 Cases

Median Percents of Errors in Estimation

Time Intervals in Seconds	1 Counting		2 Not Counting		3 Counting		4 Not Counting	
	Error	Sigma	Error	Sigma	Error	Sigma	Error	Sigma
4			29.9	2.90	26.5	3.20	51.0	2.90
6	19.3	2.40	34.3	3.50	32.0	2.75	39.0	2.75
8	26.0	2.22	34.7	2.90	25.4	1.81	31.0	1.65
11	18.6	2.40	26.3	3.50	24.4	2.75	40.5	3.20
14	25.0	2.60	32.0	2.60	26.0	1.86	48.0	2.90
18	27.0	2.22	43.7	1.86	25.0	2.05	41.0	2.60
22	28.0	2.90	35.0	2.22	31.6	2.90	42.0	2.90
27	22.7	1.86	45.3	2.60	28.0	2.75	37.0	2.40
Combined median	26.7	.75	34.2	1.02	27.7	1.03	39.3	.93
	C.R. = 5.95				C.R. = 8.03			

TABLE IV

81 Cases First Series

98 Cases Second Series

Median Percents of Errors in Estimation

Time Intervals in Seconds	1 Counting		2 Not Counting		3 Counting		4 Not Counting	
	Error	Sigma	Error	Sigma	Error	Sigma	Error	Sigma
4	26.2	2.10	28.5	2.20	26.9	4.25	29.4	3.95
6	18.4	1.95	30.7	2.85	20.0	2.70	52.0	3.70
8	25.0	1.60	28.1	2.05	27.8	2.07	29.9	5.00
11	16.9	2.00	28.5	2.73	27.1	2.50	42.5	2.00
14	12.8	1.60	27.6	2.65	27.4	2.45	75.4	5.70
18	17.7	1.30	30.0	1.92	22.6	1.67	30.0	3.40
22	12.2	1.20	33.7	2.90	22.0	2.35	38.1	2.00
27	17.3	1.96	28.6	2.40	20.0	2.47	60.2	1.98
Combined median	17.4	.81	29.0	1.05	25.4	.85	44.7	1.23
	C.R. = 8.8				C.R. = 12.8			

TABLE V

MEDIAN EXTENT AND DIRECTION OF ERRORS IN TIME ESTIMATION FOR 100 SUBJECTS

Time Intervals in Seconds	1 Counting	2 Not Counting	3 Counting	4 Not Counting
4		27.0	−14.2	−19.0
6	4.5	−19.7	−12.0	−32.3
8	2.6	−24.2	− 5.1	−15.1
11	12.5	−13.0	−14.5	−30.3
14	− 2.0	−30.0	−18.0	−33.3
18	2.1	−10.5	− 8.2	−30.0
22	− 4.8	−21.1	−16.4	−32.2
27	− 3.5	−36.2	−17.0	−28.2
Median for each series	1.6	−16.0	−13.0	−27.7

Since the experiment for counting and not counting was repeated, it is possible to determine how much consistency there is in the estimation. This was determined by combining the errors for each of the seventy subjects in group I for each of the two series of estimates when counting and when not counting. These errors were then correlated. In this correlation the averages were computed algebraically, *i.e.*, by observing the signs which indicate the direction of the error. The correlation between the errors for counting on two different tests taken about two weeks apart for Group I is .65, σ.09; between not counting it is .50, σ.10; between counting and not-counting but taken immediately after the other it is .60, σ.09 and .68, σ.07. The correlation for the 32 subjects of Group II for two successive counting series is .88, σ.05; and for a counting and non-counting series it is .83, σ.07. The correlations for Group II are higher than for Group I, probably because the subjects gave more careful attention to their estimates. In all cases there is considerable consistency in the estimates from one time to another either when counting or not counting and even between counting and non-counting. The median percents of error on the two trials for the first groups of 70 subjects when counting is 27.6, σ1.18and 26.7, σ1.00, and for not counting it is 27.6, σ1.00 and 28.2, σ.83. The critical ratios are 0 and 1.15 showing that the difference in favor of counting is very small for this group. In the second group the percents of error for counting is 28.9, σ1.75 and 24.9, σ1.38, and for not counting it is 29.1, σ1.80

and 28.3, σ1.73. The critical ratios are .08 and 1.54 which does not indicate that they are necessarily significant differences but they are consistent and in the same direction as Group I. In the third group the median percents for counting are 26.7, σ.75 and 27.7, σ1.03 and for not counting 34.2, σ1.02 and 39.3, σ.93. The critical ratios are 5.95 and 8.03 which are much larger than for the first two groups.

In Group IV the median percents for counting are 17.4, σ.81 and 25.4, σ.85 and for not counting they are 29.0, σ1.05 and 44.7, σ1.23. The critical ratios are 8.8 and 12.8 showing that the differences are unquestionably significant.

Let us now consider the main facts brought out by the study. In the first two experimental groups there is not a great difference between the errors in estimation for counting and not counting. However, in all but one part of one experiment there were somewhat more accurate estimations during counting. The errors tended to be proportional to the length of the interval. They generally ranged between 25 and 30 percent. These subjects gave very careful attention to the experiment and estimated as accurately as possible.

It is difficult to determine what cues are used for estimations without counting. It is very reasonable to assume that some kind of internal physiological process provides the cues, but as already mentioned, we have been unable to find evidence for such an explanation. Sub-vocally repeating numbers serially also gives an immediate basis for a response, that is, the subject is ready to respond with the last number he has repeated as soon as the interval is completed.

Although the mental content may be meager it may yet provide the basis for time estimation either when the subject counts or when he does not count. Some of our studies not yet published as well as those by Gullickson [7] and Harton [8] clearly indicate that when the time interval is filled with active mental processes the nature of the filling makes a great difference in the estimations. It may greatly lengthen or it may shorten the apparent length of the time.

[7] Gullickson, H. Occupation and perception of time. *J. exp. Psych.*, 1927, 10, 52–59.

[8] Harton, J. H. The influence of degree of unity of organization on the estimation of time. *J. gen. Psych.*, 1939, 21, 25–49.

In contrast to the first two groups the third and fourth groups gave only casual attention to the experiment. This was especially true when the experiment was repeated. In the first estimates while counting the errors were of about the same order as the first two groups. In fact in Group IV in the first estimates with counting the errors were smaller than in any other part of the experiment. When these subjects estimated without counting the errors were greater and the differences were statistically significant. When the experiment was repeated the estimates without counting were on the average about one-half larger than with counting. This fact is brought out in another way by the larger sigma values. There was also nearly one-third more errors in estimation of over 100 percent in the second test of the group giving least attention when they did not count. This advantage for counting when the subjects were not so attentive probably comes from several sources. Counting provides a convenient mental content for filling in the time interval. It is more constant than most other types of mental content. The motor rhythm of counting gives further clues by which the subject may tend to keep his estimates of the length of a second constant from one judgment to another.

In order to determine the general direction of errors, Table V is presented. In this table the median percents of error are arranged algebraically. This table gives the direction of the errors for the third group of one hundred subjects. Inspection of the data for the other groups indicates that the results are also similar for these groups. These results show that there is some tendency for subjects to overestimate time intervals on the first few trials. (The order of the first eight trials was 11, 6, 18, 27, 8, 22, and 14 seconds.) After the first few trials there is a general tendency to underestimate the intervals. The tendency seems to increase on successive estimates in any trial rather than, as some invesitgators have found, in relation to the length of the interval. This might be construed to indicate a tendency for an increase of tension which would probably develop during any test series and cause an over-estimation of the time, *i.e.*, to make the time appear

to be longer than it actually is. This problem deserves further study.

The Sense of Time test of the Seashore Tests of Musical Talent was also given to 52 of the first group of students and the results correlated with their time estimation scores. The correlation with the counting series was .06, σ.12 and with the non-counting series it was .19, σ.12. These results indicate that there is not a very close relationship between ability to estimate short time intervals of several seconds and the ability to discriminate the longer or shorter of two intervals of approximately one second each. Whether the low correlations are due to the difference between the length of the intervals estimated, or to the different methods of estimating, cannot be determined from this study. Probably both were contributory factors.

Most earlier studies [9] of time estimation have shown men to be more accurate in their estimations than women. The first and third groups used in the present study were separated according to sex. In both groups the women did slightly better than men, but in neither group was the difference sta-

A Comparison of Men and Women on Time Estimation Group III

	Counting	Not Counting
Men	27.8	37.03
Women	26.2	36.25

tistically significant. It is probable that formerly women were not so often called upon to estimate time, but, at least in the modern college population, women are now called upon to make as many and as accurate time judgments as men.

Although we have found that time estimation for good subjects is only slightly more accurate with counting than without, a further question arises as to whether by practice counting might not become a greater aid in estimating time. In order to test this hypothesis three supplementary experiments were conducted. One group of 16 advanced undergraduate students were given four practices in estimation on alternate days during which they were permitted to count. This was followed by three additional practices in which they

[9] Gullickson, H., *op. cit.*, and Elkine, D., De l'Orientation de l'Enfant d'Age Scholaire dans les Relations Temporelles, *J. Psychol. Nor. et Pathol.*, 1928, **28**, 425–429.

did not count and by four more trials in which they counted. The order of presentation of the intervals was varied from trial to trial and the subjects did not know the length of the intervals they were estimating. Table V presents the median percents of errors for each interval for each trial together with the means of these medians. This experiment indicates, as might be expected, that there is no improvement in estimation due to practice without knowledge of the amount and direction of the error. The extent of the errors is typical for good subjects both when they are counting and when they are not counting.

TABLE VI

SHOWING MEDIAN PERCENTS OF ERROR AND MEANS OF THESE PERCENTS FOR SIXTEEN SUBJECTS IN ELEVEN SUCCESSIVE PRACTICE PERIODS IN TIME ESTIMATION, WITHOUT KNOWLEDGE OF THE DIRECTION AND EXTENT OF ERROR

Time Intervals in Seconds	Median Percents of Error										
	Counting First Series				Not Counting			Counting Second Series			
No. of Trial.....	1	2	3	4	5	6	7	8	9	10	11
4	25	25	25	25	25	25	25	14	16	25	25
6	17	16	16	16	33	17	16	25	21	16	16
8	25	25	25	12	25	38	25	09	25	25	25
11	18	27	18	18	27	36	36	25	25	27	27
14	14	14	14	14	28	21	35	18	25	29	21
18	16	16	16	11	41	30	44	25	22	22	22
22	22	22	22	18	25	27	38	21	25	31	27
27	29	25	22	11	38	25	37	16	28	26	30
Means of the errors	20.7	21.2	19.7	15.6	29.6	28.6	32.0	19.1	23.4	25.1	24.1

Another advanced class of 26 students was divided into two equal groups and given six practice periods in time estimation. One-half estimated by counting and one-half without counting. After each estimate was recorded, the length of the interval was announced so that the subjects knew the direction and extent of their errors. In order to prevent the subjects from learning to estimate the intervals as of any standard length the intervals were varied above or below the standards used in the other experiments. Table VII gives the median percents of error for each of eight intervals for the six practice series. It is very evident from these results that knowledge of the true time of a preceding interval makes

TABLE VII

SHOWING MEDIAN PERCENTS OF ERROR AND MEANS OF THESE PERCENTS FOR 26 SUBJECTS ON SIX PRACTICE PERIODS

Half of the subjects counted and half did not count. After each estimate the subjects were told the length of the interval.

Time Intervals in Seconds	Median Percents of Error for Each Group											
	Counting Group						Not Counting Group					
No. of Trial.....	1	2	3	4	5	6	1	2	3	4	5	6
3–4	25	0	0	0	0	20	25	25	25	25	0	20
5–6	8	17	0	17	0	14	33	17	16	33	17	29
7–9	12	12	0	6	0	11	38	31	25	25	31	33
10–12	4	0	0	9	9	8	36	14	9	18	27	12
13–15	7	7	0	7	7	3	14	29	14	14	29	27
17–19	11	17	0	11	11	6	28	20	8	39	22	18
21–23	4	9	5	9	9	5	18	12	27	28	27	14
26–28	7	8	7	9	4	8	22	26	25	22	11	12
Means of the errors	9.7	9.1	1.5	8.5	5.0	9.4	27.0	21.5	18.6	25.5	20.5	18.1

it possible for subjects to become very accurate in time estimation when they are allowed to count; but it does not seem to be nearly so much help to the subjects when they do not count. In fact, without counting there is little improvement with practice and the precents of errors remain about as large as for persons estimating without knowledge of their errors. On the other hand, counting can become a very decided aid to time estimations to those who practice such estimations with a knowledge of their errors even after a very few trials. Errors are reduced on the average to between 5 and 10 percent with half a dozen practices on a series of eight estimates.

SUMMARY

In summarizing the results of this study we may say that typical college students, when estimating short intervals of a few seconds in length and when left to use their own methods, almost invariably use some form of counting. Using such methods, giving careful attention to their estimations, they, on the average, make errors of from 25 to 30 percent. These errors may either consist of overestimations or underestimations. As a rule, the first few estimations in a series are likely

to be overestimations, but after a few trials they are more likely to be underestimations.

Although most people count when permitted to do so, when they are actively engaged in making estimates they can make almost as accurate judgments without counting. Of course, we have no objective evidence, other than their own statements, that our subjects did not count when they were not supposed to count. But counting, if it did occur, was unknown to the subjects and it is unlikely that anything overt in the form of counting occurred more than occasional thoughts of larger intervals, like, "now it has been 5 seconds," '10 seconds,' and '25 seconds,' and the like. There was, of course, some kind of mental content during these estimations. Groupings of this mental content probably constitute the cues for time estimations. Through previous experience we have learned about how much mental content passes through the mind in any given length of time. Changing this mental content, we have found, changes our estimates.

When subjects give only casual attention, their estimates are almost as accurate when they count as subjects who attend closely to the interval. In this case, the counting gives the cues for estimation. When subjects giving only casual attention do not count, they have no adequate cues, and they make errors which in one part of the experiment averaged one-half larger than when they counted.

The percent of error in estimations tended to be approximately the same for the different intervals estimated. Similar percents were also obtained in some intervals of 140 seconds which were used in part of the experiment. From these facts, it seems that, although not all the earlier studies have found evidence for the fact, nevertheless the Weber-Fechner Law holds true in time estimation.

There is considerable consistency in a subject's ability to estimate time from one experiment to another whether the experiment be repeated immediately or several days later. This consistency includes both the direction and the amount of the error. There is little difference between men and women in estimating time.

Although counting is not very effective with the ordinary person in the estimating of time, it can become very effective. With only a small amount of practice errors can be reduced on the average from 25 to 30 percent to 5 or 10 percent by counting. Practice without counting does not result in such improvement. From this part of the study, it seems that the motor rhythm of counting can provide further cues for time estimation. This is especially true when one's casual attention is given to the estimation or for the highest degree of efficiency in estimation as a result of practice.

Conclusions

This study gives further information both about *how accurately* as well as *how* we estimate time. Since estimation is one very important feature in time perception, it has both practical and theoretical implications. It is not necessary to point out here the practical importance of developing methods of better time estimation.

Time and space perception are two fundamental problems in psychology as well as in many other sciences. William James and more recently the Gestalt psychologists have considered time and space to be very similar in many respects. However, there is considerable doubt about their similarity at least in genesis, accuracy, and the cues used in perception. Certainly, space perception is more fundamental, and it probably develops first in the child. It is, no doubt, more accurately perceived. In estimating space the cues used are largely kinæsthetic sensations from accommodation and convergence together with parallax, size of retinal image, etc. In time estimation we have not been able to find any such cues being used, and it is beginning to appear that time is perceived from some very different cues than those used in space perception.

Two further studies in time perception are now in progress. One deals with estimation versus reproduction of time intervals with children and adults. Another will compare body tension changes with time estimations.

(Manuscript received April 8, 1940)

CRITERIA OF READABILITY

BY MATTHEW LUCKIESH AND FRANK K. MOSS

Lighting Research Laboratory, General Electric Company, Nela Park, Cleveland

In various arts and practices concerned with vision and seeing, rate of reading has been assumed as a criterion of ease of reading or readability as the latter is influenced by the physical characteristics of the materials read or by the environment in which the reading is done. It will be noted that this restricted concept of readability does not include a consideration of possible unique characteristics of the reader which might influence his efficiency in reading. In conformity with this limited viewpoint, the present discussion of the criterion of rate of reading pertains to its application in the fields of typography and illumination rather than in the diagnosis of reading disabilities. Specifically, it has been assumed by various investigators that an improvement in visual conditions will result in an increase in the rate of reading. This is usually found to be the case. However, the differences in rate of reading materials which differ markedly in physical character are usually small and apparently unimportant from the socalled practical viewpoint. Furthermore, the relative influence of a given variable upon readability, as indicated by rate of reading, is often not even in qualitative agreement with that indicated by other criteria of readability (1).

As a by-product of a series of researches in seeing (2), involving reading as a visual task, we have accumulated a mass of data on the normal or usual rates of reading of subjects possessing normal vision and perhaps better than average ability in reading. These data, obtained under carefully controlled laboratory conditions, are presented herein as a basis for appraising the adequacy of rate of reading as a criterion of readability and, more generally, of ease of seeing. Obviously, such an appraisal requires a comparison of the results so obtained with those indicated by some other

criterion which is considered to be significant as a measure of readability. We have proposed and used (**2**) the rate of involuntary blinking while reading as such a measure in more than a score of major researches systematically arranged to test and appraise this criterion.

The significance of this criterion of ocular strain or effort is suggested by the statement of Ponder and Kennedy (**3**) "that the rate of blinking is closely related to the mental tension of the subject at the time and that in all probability the movements constitute a kind of relief mechanism whereby nervous energy otherwise unutilized passes into a highly facilitated path." Furthermore, Blount (**4**) concluded "that one of the reasons for blinking is to allow alteration to take place in the tension of the ocular muscles and thus eliminate early fatigue." It is not implied that we consider the rate of involuntary blinking to be a complete and final criterion of readability. However, it is a fact that we have used this criterion in more than a score of investigations (**5**) and have obtained, without exception, consistent and reproducible results. Some of these are presented in Fig. 1.

The present appraisal of the significance and the usefulness of rate of reading, as a criterion of readability, is based primarily upon the behavior and performance of the readers rather than upon theoretical considerations pertaining to the processes of reading. Obviously, any useful criterion must yield statistically significant differentials for visual conditions and materials usually encountered in practice. Therefore, our investigations involved the reading of such materials under usual, or practically obtainable, visual conditions which were carefully controlled and specified. All our subjects were selected at random from technical and clerical groups and were instructed to read at their usual rates and with their usual attitudes. In fact, participation in similar investigations has been a matter of routine for our subjects for many years. Thus it may be assumed that their reactions were not unduly influenced by the novelty of the experimental situations or by a specific technical interest in the researches. The reading matter may be described as informative and not of an emotional character.

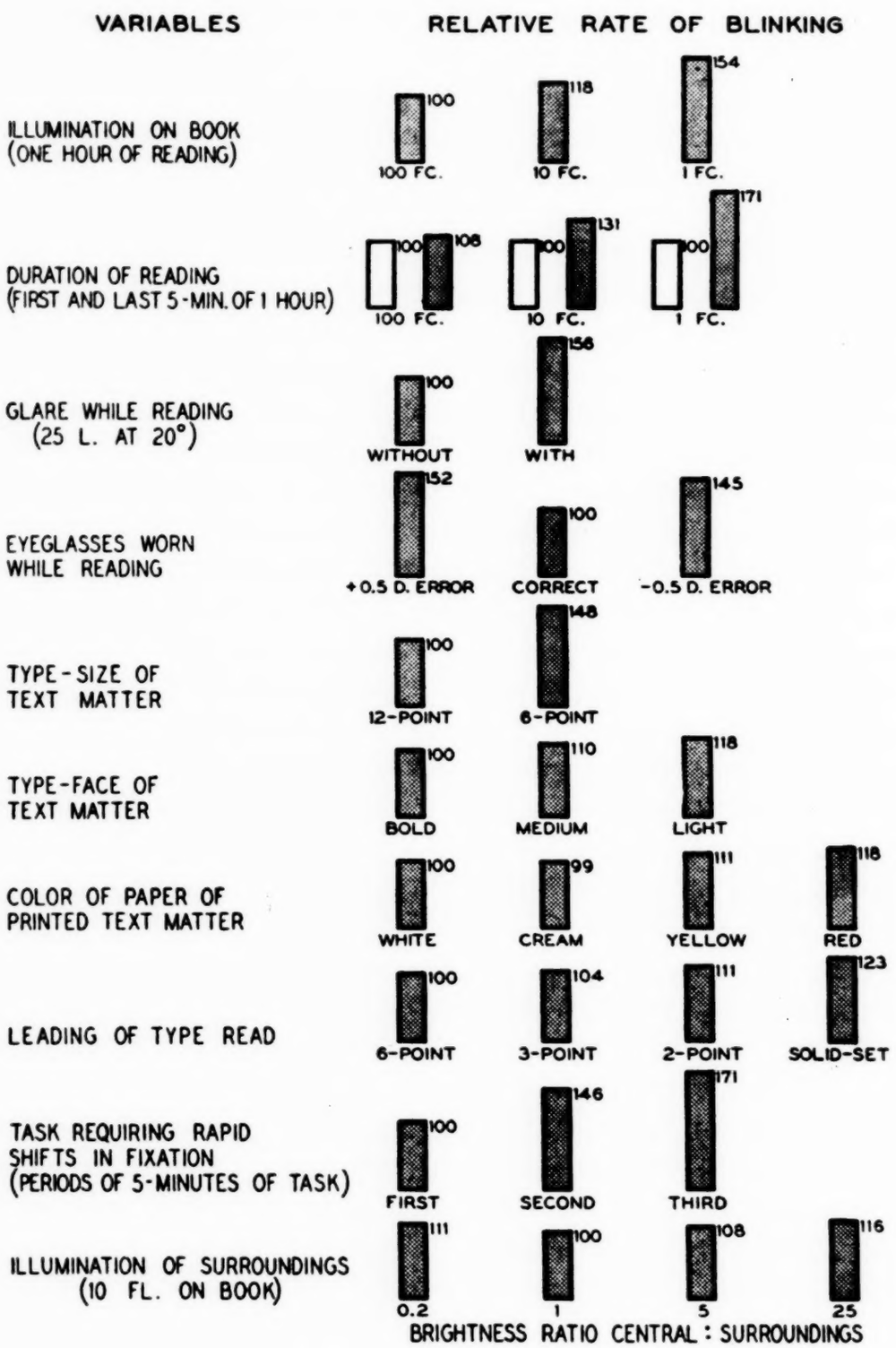


FIG. 1. The relative rates of blinking while reading as various factors which influence visibility and ease of seeing are varied individually.

Although the socalled normal rate of reading is emphasized in this discussion, it is not implied that we consider data pertaining to the maximal rate of reading for brief periods to be valueless, but rather that the influence of a given visual factor upon readability is not necessarily the same under the two conditions. For example, it should be obvious that the criterion of maximal rate of reading involves the factor of compensation to a maximal degree—a characteristic which greatly increases the complexities involved in the interpretation of the experimental results. In brief, the criterion of maximal rate of reading emphasizes the factor of 'work-output' and minimizes that of 'effort expended' in performing the task (**6**). Obviously, a consideration of the latter factor is essential in any adequate measure of readability.

Level of Illumination

The data of Tables I-*A*, I-*B* and I-*C* indicate the effect of illumination upon the normal rate of reading for periods

TABLE I-*A*

THE NORMAL RATE OF READING UNDER VARIOUS LEVELS OF DIFFUSE LIGHTING FROM 1 TO 100 FOOTCANDLES AND FOR READING-PERIODS VARYING FROM FIVE MINUTES TO ONE HOUR

The probable errors given are relative values and do not include systematic differences in rate of reading among the individual subjects.

Material: "Outline of History"—Wells. Book form.
Type: 10-pt. De Vinne; 1-pt. leading; 24 pica lines.
Tests: Two 5-minute periods by each of 10 subjects.

	1 Footcandle	10 Footcandles	100 Footcandles
Rate of reading	100	102.7	107.9
Probable error	—	1.7	1.5

TABLE I-*B*

Material: "Why We Behave Like Human Beings"—Dorsey. Book form.
Type: 12-pt. Bodoni Book; 1-pt. leading; 22 pica lines.
Tests: Eight 30-minute periods by each of 14 subjects.

	1 Footcandle	10 Footcandles	100 Footcandles
Rate of reading	100	102.9	104.7
Probable error	—	1.5	1.3

TABLE I-C

Material: "Outline of History"—Wells. Book form.
Type: 10-pt. De Vinne; 1-pt. leading; 24 pica lines.
Tests: Five 60-minute periods by each of 7 subjects.

	1 Footcandle	100 Footcandles
Rate of reading	100.0	107.0
Probable error	—	1.1

varying from five minutes to one hour. In all investigations reported in this paper, the orders of presenting the various materials or conditions to the reader were so arranged as to eliminate or minimize the influence of irrelevant factors such as progressive fatigue, learning, etc. The reading matter in all cases is printed with non-glossy "black" ink on non-glossy "white" paper.

It will be noted that a hundred-fold increase in the illumination increased the normal rate of reading from 5 to 8 percent in the situations described in Table I. However, such an increase in illumination greatly enhances the visibility of the printed-matter (7). Thus it appears that the rate of reading, although related to visibility, is not markedly influenced by the latter except under near-threshold conditions. For example, we have shown that a hundred-fold increase in illumination, from 0.1 footcandle to 10 footcandles, resulted in an increase in rate of reading of about 20 percent. On the other hand, we have found that the relative effects of other variables such as leading and brightness-contrasts in the visual field were not augmented by the expedient of determining the rate of reading under low levels of illumination.

The data of Table I also indicate that the increase in rate of reading due to increased illumination was not augmented, to any appreciable degree, by increasing the period of reading from 5 minutes to one hour. This indicates that the criterion of rate of reading is not only insensitive to changes in visibility but also is insensitive to increases in ocular fatigue since the extrinsic muscles are definitely fatigued by an hour of reading. For example, an extensive research revealed a decrement in convergence reserve of 23 percent after an hour of reading

under 1 footcandle as compared with a decrement of about 7 percent under 100 footcandles (**8**). These results suggest that the normal rate of reading is controlled largely by neural centers higher than those which govern the ocular muscles. From this viewpoint, rate of reading appears to be of limited significance as a measure of readability. In many other researches we have found that the normal rate of working, of which reading is merely a specific case, is fairly well "fixed" by many factors beyond the realm of vision and seeing.

Lighting Environment

If rate of reading is to be accepted as a criterion of readability, it should reveal the influence of such visual distractions as high brightness-contrasts within the visual field. Such contrasts are obviously unfavorable from the viewpoint of ease and comfort in seeing and in reading. Our procedures for testing the criterion of rate of reading from this viewpoint involved (1) an unrestricted field of uniform brightness surrounding the pages of an open book which was illuminated to a constant brightness of 10 footlamberts and (2) a means for varying the brightness of the surrounding field from about 0.08 footlambert to 50 footlamberts. Thus the brightness-ratio between the pages of the book and the surrounding field could be varied from 0.2 to 125. The experimental data are summarized in Table II.

TABLE II

Rates of Reading and of Blinking During 5-Minute Periods

These values involve the arithmetic mean of four tests for each individual and the geometric mean of the data from 25 subjects obtained under each condition of lighting.

Brightness-ratio (book/surrounds)	0.2	1	5	25	125
Relative rate of reading	101.6	100.0	101.2	99.6	102.7
Probable error	0.6	—	0.7	0.7	0.9
Relative rate of blinking	111.4	100.0	108.4	116.4	125.1
Probable error	1.7	—	1.4	1.4	1.8

The data of Table II indicate that the normal rate of reading was not appreciably altered by varying the brightness of the surrounding field over an extremely wide range. Hence

if it were tacitly assumed that rate of reading is an adequate criterion of ease in reading, it would follow from these results that high brightness-contrast between the central and peripheral fields is not a deleterious factor in reading. Such a conclusion is contradicted by the introspective appraisals of these conditions by the subjects. It is also in disagreement with principles established by years of research and experience in the realm of illuminating engineering (**9**).

The data obtained in this investigation permit a direct comparison of the influence of brightness-contrast (between the central and peripheral fields) upon ease of reading as the latter is indicated by the criteria of rates of reading and of involuntary blinking. It will be noted from Table II that (1) the minimum rate of blinking occurred when the brightness of the pages of the book was the same as that of its surroundings; and (2) that the rate of blinking increased markedly with increasing brightness-contrast in either direction. This result is consistent with the data obtained with various criteria pertaining to visibility (**1**), as well as with introspective appraisals of the visual conditions. However, these well-defined effects of a deleterious characteristic are not indicated by the criterion of rate of reading. Hence it seems reasonable to conclude that rate of reading is, at best, a very insensitive indicator of the subtle influences of such distractions as high brightness-contrasts within the visual field.

Typographical Factors

Leading.—It is the consensus of discerning readers that socalled solid-set type is more difficult to read than type set with greater leading or interlinear spacing. However, there is little agreement with respect to the optimum degree of leading for a given type-size, type-face, line-length, etc. Several investigators have applied the criterion of rate of reading to this phase of typography and although the individual researches have indicated optimum degrees of leading there is poor agreement among the results so obtained. In addition, the indicated optimum amount of leading is, in some cases, in such marked disagreement with the experience of

typographers that such studies have not appreciably influenced typographical practice.

We have investigated this problem (**10**) by determining (1) the normal rates of reading and (2) the frequency of involuntary blinking while reading 10-point * Textype composed in 21 pica lines in five different degrees of leading. In all investigations involving typographical variables, the text-matter consisted of test-proofs printed in black ink on white paper in galley lengths. The results are summarized in Table III.

TABLE III

These data were obtained from 30 subjects each of whom read each kind of printed matter on two different occasions. The reading was done under a constant diffuse illumination of 10 footcandles. The duration of the reading-periods was 5 minutes in all cases and the order of reading the materials was reversed in the second test-periods.

Computations of the relative rates of reading, based upon the mean of all of the data rather than upon the data for the solid-set type, yielded values almost identica to those given below. However, the probable errors were somewhat smaller and gave no indication of a systematic trend.

Points of Leading	0	1	2	3	6
Rate of reading	100.0	98.8	99.6	100.8	101.1
Probable error	—	0.7	0.9	1.0	1.2
Rate of blinking	100.0	99.6	90.3	84.1	81.2
Probable error	—	1.6	2.3	2.2	2.6

These data indicate that the normal rate of reading is substantially the same for all degrees of leading from zero to six points, although the relative probable errors of the average rates of reading are of the order of one percent. Thus if rate of reading were tacitly assumed as a criterion of readability, it would follow that leading is not an important typographical factor. Such a conclusion is contrary to the opinions of innumerable readers.

It will be noted from Table III that (1) the rate of blinking is the greatest for the solid-set and the 1-point leaded types, (2) that the rate of blinking decreases sharply as the leading is increased from 1 to 3 points, and (3) that it decreases only slightly when the leading is increased from 3 to 6 points.

* 1 point = 1/72 inch; 1 pica = 1/6 inch.

Thus it may be concluded that 3-point leading represents a practical optimum in readability of 10-point type (21 pica lines) as appraised by the criterion of the reflex blink. It is obvious from Table III that these characteristics are not revealed by the criterion of rate of reading. Thus it appears that the criterion of normal rate of reading not only does not reveal the distracting influence of lines of print adjacent to that being read, but also that it does not reveal (Table II) even the distracting influence of high brightness-contrasts within the visual field.

In view of the possibility that the criterion of *normal* rate of reading might be less likely to reveal the effects of leading than that of *maximal* rate of reading, data have also been obtained under the latter conditions. These socalled "paced" determinations of maximal rate of reading were readily made with a specially designed machine which has been described elsewhere (**9**). The results are presented in Table IV. It will be noted that the extreme variation in the

TABLE IV

AVERAGE MAXIMAL RATES OF READING 10-POINT TYPE AS OBTAINED WITH 10 SUBJECTS EACH OF WHOM READ EACH MATERIAL ON 5 DIFFERENT OCCASIONS

Points of leading	0	1	2	3	6
Rate of reading	100.0	97.2	98.8	97.6	99.3
Probable error	—	0.8	1.0	1.4	1.0

maximal rates of reading is less than 3 percent. Furthermore, the difference in the rates of reading any two materials is not significant when compared with the corresponding probable error of the difference. Thus it is evident that the relative effects of various amounts of leading are not accentuated by reading the materials at the maximal rate rather than at the socalled normal rate.

Boldness of Type-face.—The determination of the optimum degree of boldness of a family of types of a given size involves the appraisal of at least three unique variables. These are: (1) the visibility of the individual printed characters; (2) the effective amount of leading as the latter is influenced by the

type-face; and (3) the distracting influence of characters adjacent to those included in the perceptual span of the moment. Thus the appraisal of optimum boldness may constitute a more critical test of proposed criteria of readability than tests which involve but one major variable. The present study (**11**) of boldness, based upon the criteria of rates of reading and of blinking, involved four weights of Memphis type each of which was composed in 21 pica lines with two points of leading. The text was from the 'Outline of History' by Wells and the subjects were instructed to read at their normal rates. As in all researches reported herein, the order of reading the various materials was reversed in the second of the two test-periods. The data so obtained are summarized in Table V and presented graphically in Fig. 2.

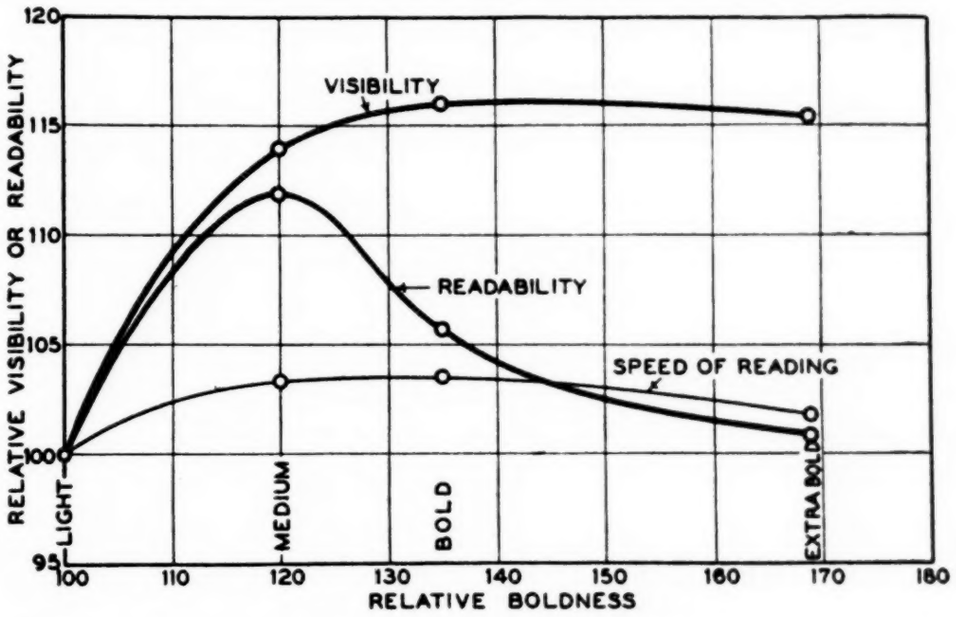


FIG. 2. In this presentation of the blink-rate data of Table V, the relationship between boldness and rate of blinking is inverted and designated as the relationship between *boldness* and *readability*. The data relating *boldness* and *visibility* were obtained with the Luckiesh-Moss Visibility Meter. The design and calibration of this instrument have been described elsewhere.[1]

These results again confirm the conclusion that the normal rate of reading is, at best, a relatively insensitive indicator of the influences of physical characteristics of the materials read. For example, it will be noted that the maximum differential in rate of reading these materials did not exceed 3.5 percent.

TABLE V

These averages were obtained from 40 subjects each of whom read each of the materials on two different occasions under a constant level of diffuse lighting of 10 footcandles. All reading periods were 5 minutes in duration.

Type-face	Light	Medium	Bold	Extra-Bold*
Rate of reading	100.0	103.2	103.5	101.8
Probable error	—	0.7	1.1	1.0
Rate of blinking	100.0	88.1	94.3	99.1
Probable error	—	1.9	2.6	3.2

* Although the Extra-Bold type is designated by the typographer as 10-point type, the actual "size" of the type, excluding ascenders or descenders, is about 10 percent larger than that of the other three types.

It is also a reasonable assumption that differentials of this order of magnitude would scarcely be considered as important or impressive from the practical viewpoint—and it is emphasized that the purpose of readability studies is to guide practice. Furthermore, the data of Table V indicate that Memphis Medium and Memphis Bold types were read at substantially the same rates. Thus these two types would be rated as alike in readability by the criterion of rate of reading. Such a rating is not in agreement with that indicated by the criterion of blinking or by the experience of typographers.

It will be noted from Table V that a maximum differential of about 12 percent in rate of blinking was observed during the reading of these materials. This is more than three times as great as that obtained in rate of normal reading. Furthermore, it will be obvious from Fig. 2 that the two criteria differ not only in sensitivity but also as indicators of relative readability.

Although the indicated differences in readability among the 10-point Memphis types appear to be real from a statistical viewpoint, the significance of these data would be greatly enhanced if approximately the same relative values were obtained under somewhat different experimental conditions. Accordingly, the investigation was repeated by a different experimenter under the following conditions:

In a different laboratory and under a level of illumination of 5 footcandles instead of 10 footcandles. Thus the environmental conditions under which the reading was done were completely altered.

The order of presenting the several weights of Memphis types was altered. In the first investigation, the orders were (*a*) Light, Medium and Bold; and (*b*) Bold, Medium and Light; and in the second, the orders were (*a*) Medium, Light and Bold and (*b*) Bold, Light and Medium. It will be noted that the positions of the type of highest readability (Memphis Medium) and of the lowest readability (Memphis Light) were transposed in the second investigation.

The measurements were made with 18 instead of 40 subjects, each of whom read each type on four occasions instead of twice as in the first investigation. Since the same materials were read, the rates of blinking obtained in the second investigation were those resulting during a re-reading of the text-matter for the second and third times.

The relative rates of blinking in the two investigations are given in Table VI.

TABLE VI

Type-face	Light	Medium	Bold
Arithmetic means			
First test	100.0	90.4	96.0
Second test	100.0	88.9	94.5
Geometric means			
First test	100.0	88.1	94.3
Second test	100.0	88.8	95.0

The agreement between the results of these two investigations emphasizes, not only from a qualitative but also from a quantitative viewpoint, the inherent significance and reliability of the criterion of involuntary blinking. Specifically, the statistical possibility that this agreement is due to chance is practically nil. Considered in conjunction with the fact that reproducible results were also obtained in repeated investigations involving other and radically different visual variables it would seem that the reliability of the blink-rate criterion has been definitely established.

Line-length.—It is conceivable that rather marked differences in rate of reading might be obtained with materials differing widely in line-length. Obviously, the muscular activities involved in reading are influenced directly by line-length; and it is possible that a particular line-length might be most favorable for the processes of reading which have

been developed by the individual. We have investigated these possibilities by determining the rate of reading, as well as the rate of blinking while reading, for materials varying in line-length from 13 to 29 picas, inclusively. These materials were printed in 10-point Textype with 2 points of leading and in black ink on non-glossy white paper. In all cases, the reading was done under a uniform diffuse illumination of 10 footcandles.

TABLE VII

THE GEOMETRIC AVERAGE RATES OF READING AND RATES OF BLINKING AS DETERMINED WITH 24 SUBJECTS

The corresponding probable errors are relative values which do not include systematic differences among individual subjects. These data represent the average results obtained in two separate investigations involving 12 subjects each. Almost identical results were obtained in the two cases.

Line-length in picas	13	17	21	25	29
Rate of reading	100.0	97.2	96.7	98.0	99.6
Probable error	—	0.9	0.9	1.2	1.4
Rate of blinking	100.0	110.4	122.0	118.2	119.5
Probable error	—	3.4	3.8	3.3	3.3

The data of Table VII summarize the results obtained with 24 adult subjects each of whom read each material on two different occasions. The order of reading these materials was reversed in the second series of tests. In all cases, the duration of the reading-periods was five minutes.

It will be noted that the maximum difference in rate of reading is less than 4 percent. Hence it appears that the criterion of rate of reading is no more sensitive to changes in line-length than it is to changes in illumination, leading, boldness, etc. This fact substantiates the theory that the normal rate of reading is limited by perceptual phenomena rather than by the physical characteristics of the visual stimuli when the reading is done under the usual supra-threshold conditions.

The data pertaining to the rate of involuntary blinking reveal relatively large differentials as compared with those involving rate of reading. Furthermore, the results obtained

with the two criteria are not even in qualitative agreement. For example, the criterion of rate of reading indicates that the readability of 29 pica lines is the same as that of 13 pica lines; while the criterion of rate of blinking indicates a marked superiority of the 13 pica lines. The latter conclusion is also supported by theoretical considerations as well as by appraisals based upon other behavioristic criteria.

Summary

Readability has been defined as an attribute of the physical characteristics of the materials read and has been appraised, on this basis, by the criteria of rate of reading and rate of involuntary blinking while reading. The results of a series of carefully controlled investigations, involving diverse visual variables, reveal that the normal rate of reading is an insensitive indicator of readability as compared with rate of blinking. Furthermore, the appraisals of readability by the two criteria are frequently in disagreement. Whether or not the criterion of involuntary blinking adequately appraises all factors involved in readability, it appears to be far more significant, from both theoretical and experimental viewpoints, than the criterion of rate of reading.

* * *

We are indebted to the Mergenthaler Linotype Company, Brooklyn, for furnishing a large amount of typographical materials specially composed and printed for us; and to S. K. Guth and A. A. Eastman for prosecuting the extensive laboratory investigations.

(Manuscript received June 22, 1940)

References

1. Luckiesh, M. and Moss, F. K., *The Science of Seeing*, D. Van Nostrand Company, New York, 1937.
2. Luckiesh, M. and Moss, F. K., The Eyelid Reflex as a Criterion of Ocular Fatigue, *J. exper. Psychol.*, 1937, **20**, 589.

Luckiesh, M. and Moss, F. K., Visibility and Readability of Print on White and Tinted Papers, *Sight-Saving Rev.*, 1938, **8**, 123.

Luckiesh, M. and Moss, F. K., Frequency of Blinking as a Clinical Criterion of Ease of Seeing, *Amer. J. Ophthal.*, 1939, **22**, 616.

Luckiesh, M. and Moss, F. K., The Visibility and Readability of Printed Matter, *J. applied Psychol.*, 1939, **23**, 645.

LUCKIESH, M. and MOSS, F. K., The Readability of Stencil Duplicated Materials, *Sight-Saving Rev.*, 1939, **9**, 295.

3. PONDER, E. and KENNEDY, W. P., On the Act of Blinking, *Quart. J. exper. Physiol.*, 1927–28, **18**, 89.
4. BLOUNT, W. P., Studies on the Movements of the Eyelids of Animals, *Quart. J. exper. Physiol.*, 1927–28, **18**, 111.
5. LUCKIESH, M. and MOSS, F. K., A Summary of Researches Involving Blink-Rate as a Criterion of Ease of Seeing, *Illumin. Eng.*, 1940, **35**, 19.
6. LUCKIESH, M. and MOSS, F. K., The Human Seeing-Machine, *Trans. Illumin. Eng. Soc.*, 1932, **27**, 699.
7. LUCKIESH, M. and MOSS, F. K., The Relative Visibility of Print in Terms of Illumination Intensity, *Sight-Saving Rev.*, 1935, **5**, 272.
8. LUCKIESH, M. and MOSS, F. K., Fatigue of Convergence Induced by Reading as a Function of Illumination Intensity, *Amer. J. Ophthal.*, 1935, **18**, 319.
9. LUCKIESH, M. and MOSS, F. K., Brightness-Contrasts in Seeing, *Trans. Illumin. Eng. Soc.*, 1939, **34**, 571.
10. LUCKIESH, M. and MOSS, F. K., Effects of Leading on Readability, *J. App. Psychol.*, 1938, **22**, 140.
11. LUCKIESH, M. and MOSS, F. K., Boldness as a Factor in Type-Design and Typography, *J. App. Psychol.*, 1940, **24**, 170.

EXPERIMENTAL STUDIES IN ROTE-LEARNING THEORY. VII. DISTRIBUTION OF PRACTICE WITH VARYING LENGTHS OF LIST

BY CARL IVER HOVLAND

Department of Psychology, Institute of Human Relations, Yale University

Introduction

Over twenty-five years ago Lyon (**8**) completed his monumental study concerning the relationship between length of material and distribution of practice. His results clearly indicate that in the learning of nonsense syllable materials the superiority of distributed over massed practice becomes progressively greater the longer the list. This general outcome is of considerable theoretical importance, but details as to the nature of the relationship are not revealed by his study. The need for additional information on this problem, particularly in connection with such analytical measures of learning as acquisition curves, serial position effects and oscillations of recall, is pointedly shown in the recent attempt of Hull, Hovland, Ross, Hall, Perkins and Fitch (**6**) to formulate a systematic theory of rote-learning. The present study was therefore designed to obtain further data concerning the relationship between length of material and the efficacy of distributed practice and to test specific predictions derived from this theoretical system.

Experimental Procedure

Three lengths of nonsense syllable lists were used for the comparison of learning by massed and distributed practice. Six different experimental programs were therefore involved. Each subject completed two cycles of the six programs, learning only a single list per day.

Use of lists nine, twelve, and fifteen syllables in length permitted comparison with results of both Robinson and Heron (**11**) and Shipley (**12**). The former used these identical lengths without a separate cue syllable; the latter employed a separate cue syllable, so that the lists he labelled eight and fourteen syllables in length actually consisted of nine and fifteen syllables, respectively. In line with recent methods of experimentation, the first syllable was also used as a cue in the present experiment. The lists would therefore be designated as eight, eleven and fourteen units in length

and will be referred to in this way throughout the paper. The lists were constructed according to the usual rules and were made up of syllables of low associational value. Systematic rotation of the series was employed as previously described (1, 5). Complete counterbalancing of practice orders was also provided.

The lists were presented by means of the Hull exposure apparatus. The general procedures and instructions used in other experiments in the present series (1, 2) were followed. Distractions were minimized by performing the experiments in a soundproof room.

During massed practice only a six-second interval elapsed between successive presentations of the lists; during distributed practice a two-minute rest pause was given between trials. To prevent rehearsal during this period, the subjects were required to name colors automatically presented on the memory drum. A two-second rate of syllable presentation was employed throughout. Learning was carried to the criterion of one perfect recitation for all lists. The individual responses throughout the learning were recorded on specially prepared mimeographed forms.

Thirty-two Yale College students served as subjects. They had not been previously employed in memory experimentation and were consequently given a whole cycle of practice, with one session devoted to each of the six variations of length and distribution. The practice lists used for the purpose contained words of low or medium associational value.

Results and Discussion

Data are presented in Table 1 concerning the number of trials required to reach the criterion of one perfect recitation

TABLE 1

Mean Numbers of Trials Required to Learn Three Lengths of List by Massed and Distributed Practice to the Criterion of One Perfect Recitation (each Mean Based upon 2 Values for each of 32 Subjects)

Method of Practice		Length of List (in syllables)		
		8	11	14
Massed	M	7.00	11.64	17.14
	S.E.	0.28	0.68	0.80
Distributed	M	5.93	9.43	12.14
	S.E.	0.36	0.59	0.76
Differences	M	1.07	2.21	5.00
	S.E.	0.33	0.62	0.97
C.R.		3.24	3.56	5.16

in the learning of each of the three lengths of list by massed and distributed practice. Distributed practice is superior to massed for all lengths. Further, the reduction in number of trials as a result of distributed practice becomes progressively greater the longer the list. On this point the present results support those of Lyon (8), and confirm Theorem XV in the

monograph by Hull, Hovland, Ross, Hall, Perkins and Fitch (**6**): *In the learning of rote series, the advantage of well distributed practice over massed practice, in terms of repetitions, is a concave function of the mean number of repetitions required for learning by massed practice* (p. 132).

Acquisition curves for the lists of various lengths learned by the two methods of distribution are graphically shown in Fig. 1. The results are plotted in terms of the number of

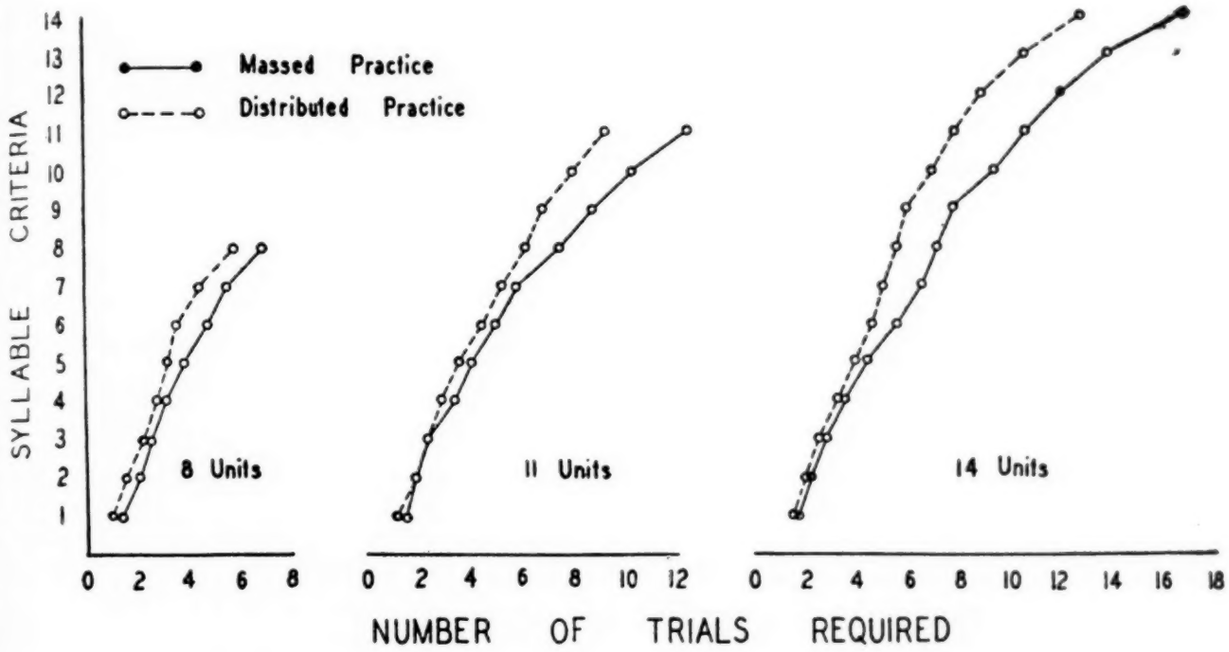


FIG. 1. Graph showing mean number of trials required to reach successive criteria of performance by massed and distributed practice for three lengths of list. Each point is based upon the mean of 64 learning scores (two from each of 32 subjects).

repetitions required to reach successive criteria of performance. The results indicate a relatively linear rate of acquisition even with massed practice in the case of the shorter lists, but quite pronounced negative acceleration in the learning of the longer lists (*Cf.* Kjerstad (**7**)). Curves of acquisition by distributed practice are more linear throughout. Departure from linearity was computed in terms of the average size of the squares of the deviations between the values actually observed and those computed from a straight line fitted to the data by the method of least squares. The average sizes of these squared deviations are presented in Table 2 for the three

TABLE 2

AVERAGE VALUES OF SQUARES OF DEVIATIONS OF OBSERVED VALUES FROM STRAIGHT LINE OF BEST FIT FOR THREE LENGTHS OF LIST LEARNED BY MASSED AND DISTRIBUTED PRACTICE $\left(\frac{\Sigma D^2}{N}\right)$

Method of Practice	Length of List (in syllables)		
	8	11	14
Massed	.11	.27	.85
Distributed	.08	.14	.29

lengths of lists learned by massed and distributed practice. It will be seen that the deviation from linearity is greater the longer the list, and is less with distributed than with massed practice.

An interesting comparison can be made between the number of repetitions per syllable in the present experiment and in that of Robinson and Heron (11). The data from their first two cycles can be compared with the present results concerning learning by massed practice. The average numbers of repetitions per syllable in the Robinson-Heron experiment were .86, .98, and 1.09 for the nine, twelve, and fifteen units' lengths, respectively; in the present experiment the corresponding results were .78, .97, and 1.14, when the lengths were computed in a manner comparable to theirs. The close agreement is the more surprising because in their study the letters in the syllables were individually pronounced, while in the present experiment the syllables were recited as units. The corresponding numbers of repetitions per syllable for distributed practice learning are .66, .78, and .81 for the nine, twelve, and fifteen unit lists, respectively, computed in the manner of Robinson and Heron. The rate of increase in difficulty with increased length is therefore less with distributed than with massed practice.

Serial Position Effects.—Figure 2 presents the number of failures in the various syllable positions during the learning of the three lengths by massed and distributed practice.

The greater difficulty of the central syllables as compared with the ends is clearly shown for all the curves, the longer the

list the greater the excess of difficulty in the middle. Likewise, as in most previous experiments, the difficulty of the posterior syllable is greater than that of the anterior. This

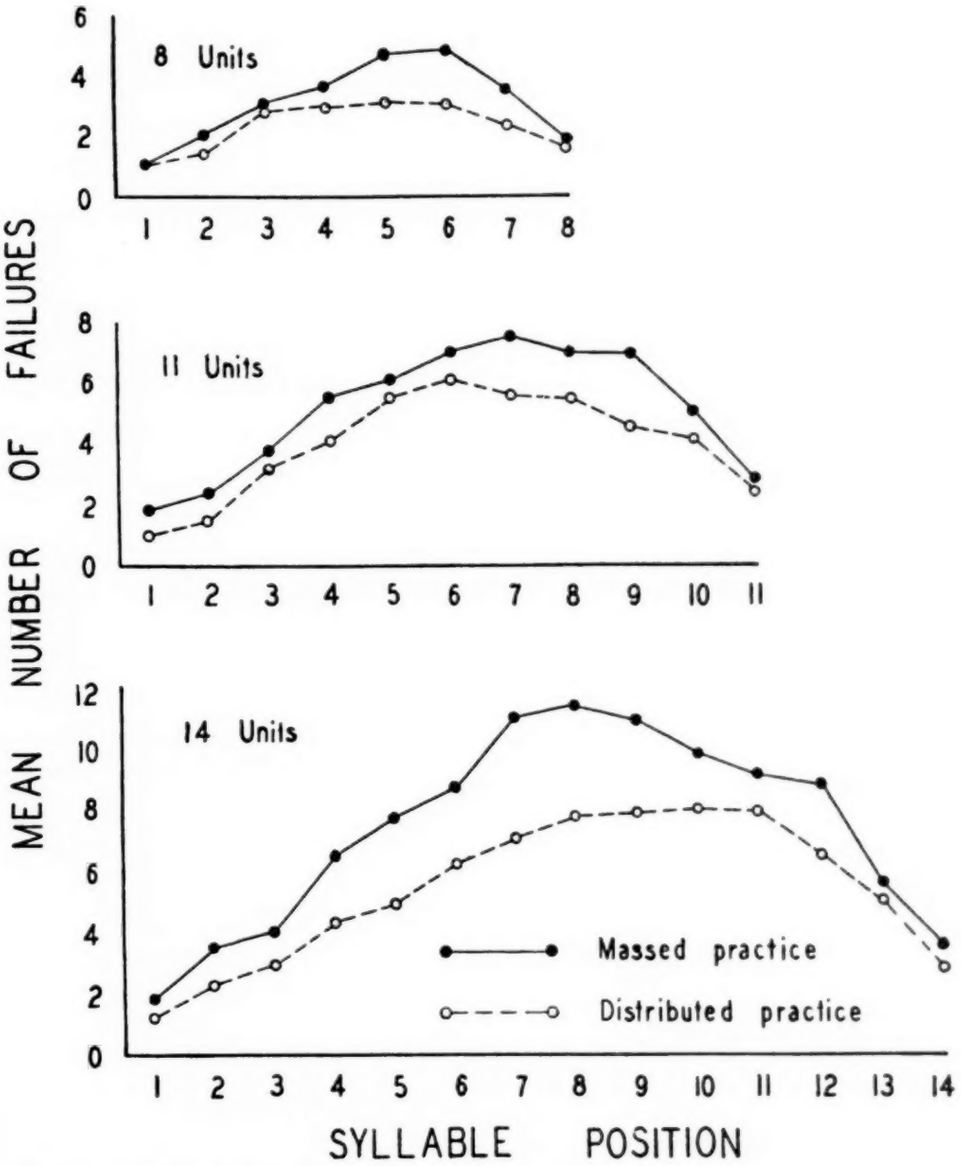


FIG. 2. Composite curves showing mean number of failures at various syllables during learning of three lengths of list by massed and distributed practice. Each point is based upon the mean of 64 learning scores.

is true for all three curves of learning by massed practice presented in Fig. 2. The probabilities of obtaining these differences by chance are all less than .01. The ratios of the number of failures at the posterior syllable divided by the number at the anterior, range from 1.64 for the eight unit to

1.87 for the fourteen unit lengths. The greater difficulty of the central syllable as compared with the ends, and the anterior as compared with the posterior syllable, also obtains with distributed practice.

From Fig. 2 it will be clearly seen that the advantage of distributed practice (the difference between the number of failures during distributed practice and the number during massed) tends to be greater in the central syllable positions than at the ends for all three lengths of list, but no statistically significant differences were obtained in the advantage of distributed practice at the anterior as compared to the posterior syllables. Patten (**9**), using a sixteen unit length list, found the advantage of distributed practice greater at the posterior than at the anterior syllable position, but it is not clear whether the difference is significant.

TABLE 3

MEAN NUMBER OF FAILURES IN VARIOUS SYLLABLE POSITIONS DURING LEARNING BY MASSED AND DISTRIBUTED PRACTICE WITH THREE LENGTHS OF LIST

Syllable Position	8 Unit		11 Unit		14 Unit	
	Massed	Distributed	Massed	Distributed	Massed	Distributed
First	1.17	1.01	1.80	1.07	1.98	1.24
Second	2.14	1.46	2.36	1.50	3.64	2.35
Third	3.14	2.80	3.86	3.28	4.14	3.07
Fourth	3.60	2.95	5.51	4.14	6.58	4.36
4th from posterior	4.69	3.14	7.07	5.52	9.27	7.01
3rd from posterior	4.82	3.06	7.01	4.53	8.92	6.64
2nd from posterior	3.50	2.36	5.14	4.21	5.72	5.07
Posterior	1.92	1.71	2.92	2.10	3.71	2.98

For convenience in making cross-comparisons, Table 3 presents the number of failures made in corresponding syllable positions for the three lengths. It will readily be seen that with increased length the number of failures goes up in each serial position. Thus, for syllable 1, the mean numbers of repetitions required are 1.17 for an eight unit list, 1.80 for an eleven unit and 1.98 for a fourteen unit list. Similarly, for syllable 4 the corresponding figures are 3.60, 5.51 and 6.58.

A tendency is also indicated for the increase in failures with increased length to be most rapid in the central region,

less at the posterior end and least at the anterior. Thus, in the anterior syllable position there is an average increase of 0.40 failures with each three syllable increase in length. The corresponding increases are 0.90 failures at the posterior syllable and 2.29 failures at the fourth syllable from the end.

Similarly, there is a pronounced tendency for the advantage of distributed practice to increase more rapidly with increased length in the central syllable positions than at the ends. The average increase in advantage of distributed practice with each three syllable increase in length is 0.29 failures at the anterior syllable. The corresponding increases are 0.78 for the fourth syllable position and 0.26 for the posterior syllable. The differences in the rates of change at the anterior and posterior syllables will be seen to be slight.

Oscillations at the Threshold of Recall.—Hull (**4**) has shown the usefulness of computing the span between the occasion on which the first success is achieved and that on which the last failure is made for the various syllable positions. In Table 4 results on this point are presented. For each position the mean number of failures preceding the intitial success and up to and including the last failure are given. If no oscillations in recall are present, these two values coincide. If, however, the subject is able to recall a syllable on one occasion but not on a later trial, the oscillation shows itself in a discrepancy between the failures preceding the first success and the number up to and including the last failure.

In the present experiment, as in previous studies (Hull (**5**), Hovland (**3**) and Shipley (**12**)), a characteristic serial position effect is obtained with respect to initial success and last failure (Table 4). This is brought out forcibly in Table 5, where the mean numbers of repetitions preceding the first success and last failure in the first, last and middle positions are given. The greatest numbers of failures regularly occur in the central syllable positions. This effect becomes progressively greater the longer the list. It will also be noted that the last syllable is preceded by more failures than the first for every list with the exception of the eight unit length learned by massed practice.

TABLE 4

MEAN NUMBER OF REPETITIONS (*a*) PRECEDING FIRST SUCCESS AND (*b*) UP TO AND INCLUDING LAST FAILURE IN VARIOUS SYLLABLE POSITIONS DURING MASSED AND DISTRIBUTED PRACTICE

Each value is mean of 64 scores—two upon each of 32 subjects

Mean No. of Repetitions	Length of List	Syllable Position													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Massed Practice															
Preceding First Success	8	0.94	1.50	2.93	2.92	4.38	4.01	3.22	1.10						
	11	1.01	0.78	2.50	4.56	4.92	5.80	6.24	5.53	5.76	3.92	2.22			
	14	1.03	2.16	2.86	5.43	6.02	6.40	7.60	9.20	7.65	7.93	6.80	6.36	3.65	2.86
Up to and including Last Failure	8	2.37	3.72	4.71	4.80	4.89	4.90	3.79	2.28						
	11	3.36	4.82	6.78	6.57	7.57	8.38	8.60	8.40	8.65	7.57	4.71			
	14	4.28	6.14	8.21	10.37	10.04	12.70	14.02	13.49	14.48	13.30	12.60	11.79	9.72	5.93
Distributed Practice															
Preceding First Success	8	0.71	1.07	2.35	2.14	2.64	2.85	2.30	1.42						
	11	0.78	1.14	2.57	3.14	4.57	5.78	4.14	5.64	4.07	3.14	2.28			
	14	0.71	1.85	2.07	3.85	4.42	5.07	6.21	5.14	5.21	5.85	6.35	5.50	4.14	2.78
Up to and including Last Failure	8	1.50	1.93	3.22	3.36	3.78	3.78	2.64	2.07						
	11	2.64	1.85	4.42	6.71	7.07	6.85	7.35	6.50	6.14	6.78	2.92			
	14	2.50	2.92	3.78	6.00	6.28	8.71	9.50	9.14	8.78	8.92	8.71	8.00	7.14	3.71

TABLE 5

MEAN NUMBER OF REPETITIONS PRECEDING *First Success* AND UP TO AND INCLUDING *Last Failure* IN VARIOUS POSITIONS OF LIST (DATA FROM TABLE 4)

Length of List	First Success			Last Failure		
	Syllable Positions			Syllable Positions		
	First 2	Middle Group	Last 2	First 2	Middle Group	Last 2
Massed Practice						
8 unit	1.22	3.65	2.16	3.04	4.84	3.03
11 unit	0.89	5.65	3.07	4.09	8.18	6.14
14 unit	1.58	8.40	3.25	5.22	13.75	7.82
Distributed Practice						
8 unit	.89	2.39	1.86	1.71	3.57	2.35
11 unit	.96	4.83	2.71	2.24	7.09	4.85
14 unit	1.28	5.67	3.46	2.71	9.32	5.42

From Tables 4 and 5 it can be seen that the longer the list the greater the number of repetitions preceding the first success and up to and including the last failure in each position in the series. Thus, for example, there are .94, 1.01 and 1.03 repetitions preceding the first success for the eight, eleven and fourteen syllable lists, respectively, in the anterior syllable position, and 2.92, 4.56 and 5.43 for the fourth syllable. The numbers of repetitions up to and including the last failure in the anterior position are 2.37, 3.36 and 4.28 for the three lengths, and 4.80, 6.57 and 10.37 in the fourth syllable position. The average number of repetitions preceding the first success increases from 2.62 for the eight unit list to 3.93 for the eleven and 5.42 for the fourteen unit length.

The above results suggest differences in the rate of increase in number of repetitions with increased length for the various syllable positions. These differences can easily be demonstrated. For example, the average increases in number of repetitions preceding first success with a three syllable increase in length are 0.04 at the anterior syllable, 1.25 at the fourth syllable and 0.88 at the posterior end (learning by massed practice). As predicted by the theory (**6**), these increases are

least at the extremes and most rapid in the central positions, and more marked at the posterior than at the anterior syllable. Similarly, the average increases in number of repetitions up to and including the last failure are 0.95 at the anterior syllable, 1.82 at the posterior, and 2.78 at the fourth syllable position.

From Table 4 may also be derived the fact that with increased length the number of repetitions preceding the first success increases at a less rapid rate than those preceding the last failure. The average increase in mean number of repetitions (for all syllables) with a three syllable increase in length, is 1.40 preceding the first success, but 3.28 up to and including the last failure.

Distributed practice results in a decreased number of repetions preceding the first success and up to and including the last failure for all lengths of list, the decrease being more marked the longer the list, and greater in terms of last failure than first success. The rate of increase in number of repetitions with increased length is smaller with distributed practice (average changes with a three syllable increase in length are 1.15 repetitions preceding first success and 1.97 up to and including last failure, as compared with corresponding values of 1.40 and 3.30 for massed practice).

Table 6 gives the average number of repetitions in the first two, last two, and central syllable positions between the first success and last failure. The longer the list the greater the span between first success and last failure in each syllable position. The average span for the eight unit list is 1.31 repetitions, that for the eleven unit, 2.92, and for the fourteen unit, 5.08 (learning by massed practice). On theoretical grounds it would also be expected that the greatest spread between first success and last failure would be obtained in the central portion. This expectation is fulfilled in the case of distributed practice. With massed practice, however, only the longest list has the greatest spread in the middle. The shortest list shows the greatest spread in the anterior region, and the middle length at the ends. In this respect the present data parallel those of Shipley (**12**). These results do not neces-

TABLE 6

MEAN NUMBER OF REPETITIONS BETWEEN FIRST SUCCESS AND LAST FAILURE IN VARIOUS POSITIONS OF LIST (DATA FROM TABLE 5)

Length of List	Syllable Positions		
	First 2	Middle Group	Last 2
Massed Practice			
8 unit	1.82	1.19	.87
11 unit	3.20	2.53	3.07
14 unit	3.64	5.35	4.57
Distributed Practice			
8 unit	.82	1.18	.49
11 unit	1.28	2.26	2.14
14 unit	1.43	3.65	1.96

sarily contradict the theoretical predictions (**6**, p. 181), which are based upon the supposition that the learning will be carried to a point of complete mastery, with a criterion of an infinite number of perfect recitations. In the present experiment the criterion was only a single perfect trial. The lower criterion obviously introduces an artificial 'ceiling,' since the trial on which the most difficult syllables are finally learned tends to be the last trial; consequently the first success and last failure tend to coincide at the points of maximum difficulty. With further trials it would be expected that pronounced oscillations in recall would be found in the most difficult syllable positions and the predictions might thus be supported.

With the shortest list the span between the first success and last failure is greater at the anterior than at the posterior end; with the longest list this relationship is reversed (Table 6). In the case of the middle length the spans are about equal. This outcome is in agreement with the results of Shipley (**12**) and Hull (**5**). The former found the same type of discrepancy between long and short lists; the latter obtained the predicted greater span at the posterior end with a long list (16 syllables).

The span between the first success and last failure is uniformly greater with massed practice than with distributed, the longer the list the greater the difference. The average

increase in number of repetitions between first success and last failure with a three syllable increase in length is 1.88 with massed practice, only 0.82 with distributed practice.

Summary and Conclusions

An analysis was made of the learning of three lengths of nonsense syllable lists by massed and by distributed practice to a criterion of one perfect recitation. Twelve lists, two of each length and method of practice, were learned by thirty-two college student subjects at the rate of one per day. With massed practice only six seconds elapsed between successive trials; with distributed practice two-minute rest pauses were interpolated between trials. The syllables were automatically presented at a two-second rate.

Records were obtained of the number of trials required for learning, the errors in various syllable positions and the number of presentations preceding the first success and last failure in each position. The following results were found:

1. Distributed practice was superior to massed for all lengths, and the reduction in number of trials as a result of distribution became progressively greater the longer the list.

2. The longer the list the greater the negative acceleration in the curves of acquisition. Departure from linearity was more pronounced with massed than with distributed practice.

3. Despite the differences in procedure, the numbers of repetitions per syllable with massed practice were strikingly similar to those obtained by Robinson and Heron.

4. The customary greater difficulty of the central portions of the list as compared with the ends, and of the posterior syllable as compared with the anterior, were obtained for all lengths of list.

5. The efficacy of distributed practice over massed was significantly greater in the central portions than at the extremes for all lengths of list studied.

6. The advantage of distributed practice over massed was not significantly different at the anterior and posterior syllable positions.

7. With increased length of list there was an increase in

number of failures in each serial position. The rate of increase of difficulty with increased length was more rapid in the central portion of the list than at the ends.

8. The advantage of distributed practice over massed increased more rapidly with increased length in the central syllable positions than at the ends. No significant differences in rate of increase at the anterior and posterior syllable positions were obtained.

9. The numbers of trials preceding the first success and up to and including the last failure were uniformly greater at the center of the lists than at the ends.

10. With one exception more repetitions preceded the first success and last failure in the posterior than in the anterior syllable positions.

11. The longer the list, the greater the number of repetitions preceding the first success and last failure in each syllable position.

12. The rate of increase in number of repetitions preceding the first success and last failure as a result of increased length of list is most rapid for the central syllables and least rapid for those at the ends.

13. The longer the list the greater the spread between the first success and last failure.

14. With the shortest list the narrowest span between the first success and last failure was in the posterior region. With the medium length list the narrowest spread was in the central region. Only with the longest list, the fourteen unit, was the greatest spread in the central region, as predicted on the theoretical grounds. The present data are not crucial, however, since the theory demands a criterion of a large number of consecutive perfect trials, and in the present experiment only a single perfect recitation was required. Under these conditions the first success and last failure tend to coincide for the most difficult syllables, since when these are correctly recited the learning is terminated. Hence an artificial 'ceiling' is introduced.

15. The span between the first success and last failure is greater at the anterior than at the posterior end with the

shortest list; with the longest the greater span is at the posterior end. In the case of the medium length list the spans at the two ends are approximately equal.

16. With distributed practice there are fewer repetitions in each syllable position preceding the first success and last failure than with massed practice. The rate of increase in repetitions preceding first success and last failure with increased length is less rapid with distributed than with massed practice.

17. The span between the first success and last failure is greater with massed than with distributed practice. The rate of increased span with increased length is also greater with massed than with distributed practice.

(Manuscript received April 2, 1940)

References

1. Hovland, C. I., Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice, *J. exp. Psychol.*, 1938, **22**, 201–224.
2. Hovland, C. I., Experimental studies in rote-learning theory. II. Reminiscence with varying speeds of syllable presentation, *J. exp. Psychol.*, 1938, **22**, 338–353.
3. Hovland, C. I., Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation, *J. exp. Psychol.*, 1938, **23**, 172–190.
4. Hull, C. L., The formation and retention of associations among the insane, *Amer. J. Psychol.*, 1917, **28**, 419–435.
5. Hull, C. L., The influence of caffeine and other factors on certain phenomena of rote learning, *J. gen. Psychol.*, 1935, **13**, 249–274.
6. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., and Fitch, F. M., *Mathematico-deductive theory of rote learning—a study in scientific methodology*, New Haven: Yale Univ. Press, 1940.
7. Kjerstad, C. L., The form of the learning curves for memory, *Psychol. Monogr.*, 1919, **26**, No. 116.
8. Lyon, D. O., The relation of length of material to time taken for learning, and the optimum distribution of time, *J. educ. Psychol.*, 1914, **5**, 1–9, 85–91, 155–163.
9. Patten, E. F., The influence of distribution of repetitions on certain rote-learning phenomena, *J. Psychol.*, 1938, **5**, 359–374.
10. Raffel, G., Two determinants of the effect of primacy, *Amer. J. Psychol.*, 1936, **48**, 654–657.
11. Robinson, E. S., and Heron, W. T., Results of variations in length of memorized material, *J. exp. Psychol.*, 1922, **5**, 428–448.
12. Shipley, W. C., The effect of a short rest pause on retention in rote series of different lengths, *J. gen. Psychol.*, 1939, **21**, 99–117.

TIME ERRORS IN THE DISCRIMINATION OF COLOR MASS BY THE RANKING METHOD

BY B. R. PHILIP

Fordham University

The apparent bias in judgment when S attempts to discriminate sensory material, dependent upon the order of, or interval between the presentation of the standard as compared with that of the comparison stimulus, is called the time error. It is not due to a time function alone, for similar errors have been found when absolute comparisons are made, and even when there is no temporal sequence as all the material may be presented simultaneously. Trivial as the errors are in themselves they are of importance in the study of the process of judgment, for they indicate that the process is one of intricate complexity. On this point all experimenters are agreed no matter what type of explanation they advance to account for these errors. The various theories which have been adequately summarized by previous workers (**1**, **2**, **3**), are not in such great conflict as may seem at first blush. Possibly they emphasize some factor that is especially important in, though specific to, the type of judgment under investigation, and hence they may not necessarily be general explanations.

With the object of investigating one aspect of this phenomenon the following study is reported, and a tentative explanation is advanced of the relationship between the time error and the skewness of the distribution of judgments.

Procedure

The experimental material has been previously described (**4**). Briefly, there were 32 Ss, all women, who ranked 6 sets of color combinations, 12 cards to a set, for predominance of color mass. To every S 25 trials of the entire test were given, a total of 57,600 judgments. The material was spread before Ss, hence both relative and absolute judgments were made, with a predominance of the latter type of judgment toward the end of the task. Two scoring methods were utilized, the frequency of the errors weighted according to their displacement from the correct position, Method I, and a tabulation of the frequencies of allocations of every card to all serial positions,

Method II. The latter method is the preferred psychophysical one, and was used in this study, as it permitted calculation of points of subjective equality, the time errors, and the skewness and the standard deviations of the data.

Owing to the asymmetry of the distributions the time errors were calculated from the medians rather than from the means. Here again we have a choice of two methods of calculating the time errors: the difference between the point of subjective equality and the standard, and what has been called the D percent, that is, the percentage of the difference of greater or less judgments over the total number of judgments. It is the method commonly used by Needham (5), Pratt (6), and others, and though briefer, it is directly related to the former method. Owing to the high percentage of correct judgments in this study, the determination of the median, from which the point of subjective equality is calculated, is largely dependent upon the percentage of the difference of the greater and lesser judgments. The first or standard method has been chiefly used in this study, though the other method is also presented for comparison purposes.

It must be emphasized that the time error is not a measure of the accuracy of the judgment, *i.e.*, of the frequency of correct judgments, but rather it is a measure of the degree to which on the average the card is over- or underestimated. In other terms the time error gives us a clue to the apparent scale value of the stimulus in the opinion of S, *i.e.*, it enables us to form an idea of the distortion that the actual scale of standards undergoes. To phrase the idea in other terms, the time error deals with the wrong judgments, determining how they are distributed with reference to the total scale. The previous study was concerned with the ability of Ss to discriminate color mass according to serial position, as determined by the percentage of their correct judgments at each serial position.

For each color combination a table similar to Table 1 was drawn up to show the distribution of judgments according to serial position. Table 1, which was also presented in the previous study, gathers all these data together, for the distributions are very similar no matter what the color combination. It is obvious that the distributions show marked skewness and that a normal curve is approximated only for the distributions of cards 21 and 20. The truncations of the distributions owing to the physical limitations of the scale are not the only factor causing the skewness. Particularly it may be noted the center of the series, the 18 card, shows a definite truncation of judgments, as was noted in the previous paper (4), in spite

RESULTS

Distribution of Judgments

TABLE 1

FREQUENCY DISTRIBUTION OF STIMULI ACCORDING TO SERIAL POSITION (N = 57,600)

Stimuli	Position											
	23	22	21	20	19	18	18	19	20	21	22	23
23	3903	449	195	146	76	19	9	1		1		1
22	435	2931	706	455	177	55	20	13	3	1	4	
21	210	714	2503	838	343	100	49	22	10	6	3	2
20	106	373	808	2532	614	198	95	40	11	13	7	3
19	67	191	368	506	2850	494	150	111	21	21	17	4
18	53	92	133	209	463	3021	544	143	55	38	36	13
18	9	25	46	59	118	561	2990	554	200	123	79	36
19	8	11	20	29	93	201	519	2880	472	310	194	63
20	3	6	12	13	27	83	220	601	2484	823	437	91
21	2	2	6	9	24	43	101	251	880	2558	714	210
22	2	4	3	3	9	20	72	138	500	690	2867	492
23	2	2		1	6	5	31	46	164	216	442	3885

of the fact that there are no physical limitations to be found there. What does cause this effect is the method of giving the instructions which required the setting out of the cards in two rows, according to the predominating color. Thus an attitude was imparted to S whereby he had two tasks or gestalten in mind, division of the total series according to color, and serial ranking in each division according to the predominance of hue. How strong this attitude was may be seen from the abrupt cessation of judgments within each half of the color division.

From the data of the distribution tables the median rather than the mean was used to determine the points of subjective equality. The difference between the values of these points and the standards for each serial position is taken as the time error. In addition, the means and standard deviations of the distributions were determined, and the Pearsonian measure of skewness, $\frac{3(M - Md)}{\sigma \text{ dist.}}$, was calculated for each color combination and for each quartile of the total groups of Ss. The placing of S in a quartile was determined from the total error scores (Method I), on all six color combinations. In Tables

TABLE 2

TIME ERRORS ACCORDING TO SERIAL POSITION FOR EACH COLOR COMBINATION

Color	Position											
	23	22	21	20	19	18	18	19	20	21	22	23
BP	−.112	−.231	−.142	.056	.140	.093	.192	.058	.110	−.100	−.224	−.125
RG	−.095	−.163	−.058	.123	.028	.061	.110	.089	.047	−.126	−.109	−.117
BR	−.114	−.107	−.083	.070	.071	.077	.068	.012	.137	−.036	−.121	−.099
GB	−.150	−.181	−.079	.080	.007	.148	.102	.072	.030	−.106	−.200	−.132
PR	−.110	−.172	−.090	.012	.017	.105	.112	−.060	.048	−.082	−.137	−.097
GP	−.113	−.175	−.076	.025	.088	.091	.070	−.003	.117	−.021	−.224	−.136
Aver.	−.116	−.172	−.088	.061	.059	.096	.109	.028	.081	−.074	−.169	−.118

TABLE 3

SKEWNESS OF THE DISTRIBUTION CURVES FOR EACH COLOR COMBINATION

Color	Position											
	23	22	21	20	19	18	18	19	20	21	22	23
BP	−.774	−.957	−.343	−.008	.557	.516	.497	.333	.012	−.433	−.814	−.908
RG	−.726	−.671	−.178	.327	.356	.292	.642	.533	.086	−.357	−.597	−.763
BR	−.795	−.660	−.330	.087	.380	.564	.370	.666	.221	−.135	−.662	−.770
GB	−.980	−.766	−.374	.071	.303	.585	.475	.580	.025	−.364	−.731	−.796
PR	−.752	−.699	−.397	−.142	.305	.556	.426	.042	.029	−.331	−.690	−.727
GP	−.834	−.854	−.406	.034	.352	.509	.250	.362	.156	−.218	−.677	−.826
Aver.	−.810	−.768	−.336	.062	.376	.504	.443	.419	.088	−.306	−.695	−.798

TABLE 4

STANDARD DEVIATIONS OF THE DISTRIBUTION CURVES FOR EACH COLOR COMBINATION

Color	Position											
	23	22	21	20	19	18	18	19	20	21	22	23
BP	.922	1.114	1.156	1.262	1.189	1.413	1.220	1.332	1.279	1.168	1.343	1.285
RG	.720	.900	.946	.946	.936	.812	1.090	1.100	.993	1.109	1.127	.798
BR	.812	.996	.999	1.024	1.035	1.001	.876	.974	1.006	.891	.930	.760
GB	1.024	1.127	1.220	1.278	1.295	1.269	1.245	1.185	1.217	1.152	1.241	.994
PR	.922	1.028	.984	1.126	1.220	1.155	1.044	1.140	1.028	1.143	1.041	.811
GP	.925	1.168	1.187	1.147	1.238	1.253	1.095	1.211	1.214	.922	1.144	1.253
Aver.	.888	1.056	1.082	1.131	1.152	1.151	1.095	1.157	1.123	1.064	1.138	.984

2, 3, and 4 are listed the time errors, the skewness and the standard deviations of the distributions for each color combination. The time errors in Table 2 when averaged from the six color combinations yield the values given at the bottom of the table. They are in remarkably close agreement with the values of the time errors obtained from the total distribution table (Table 1), differing only in the third significant figure.

Time Error and Skewness

From the tables it is obvious that neither color combination nor color is of importance in determining the time error. In the previous paper it was shown that the difficulty of discrimination, as measured by error frequency, is partly dependent upon color combination, and that all colors are not

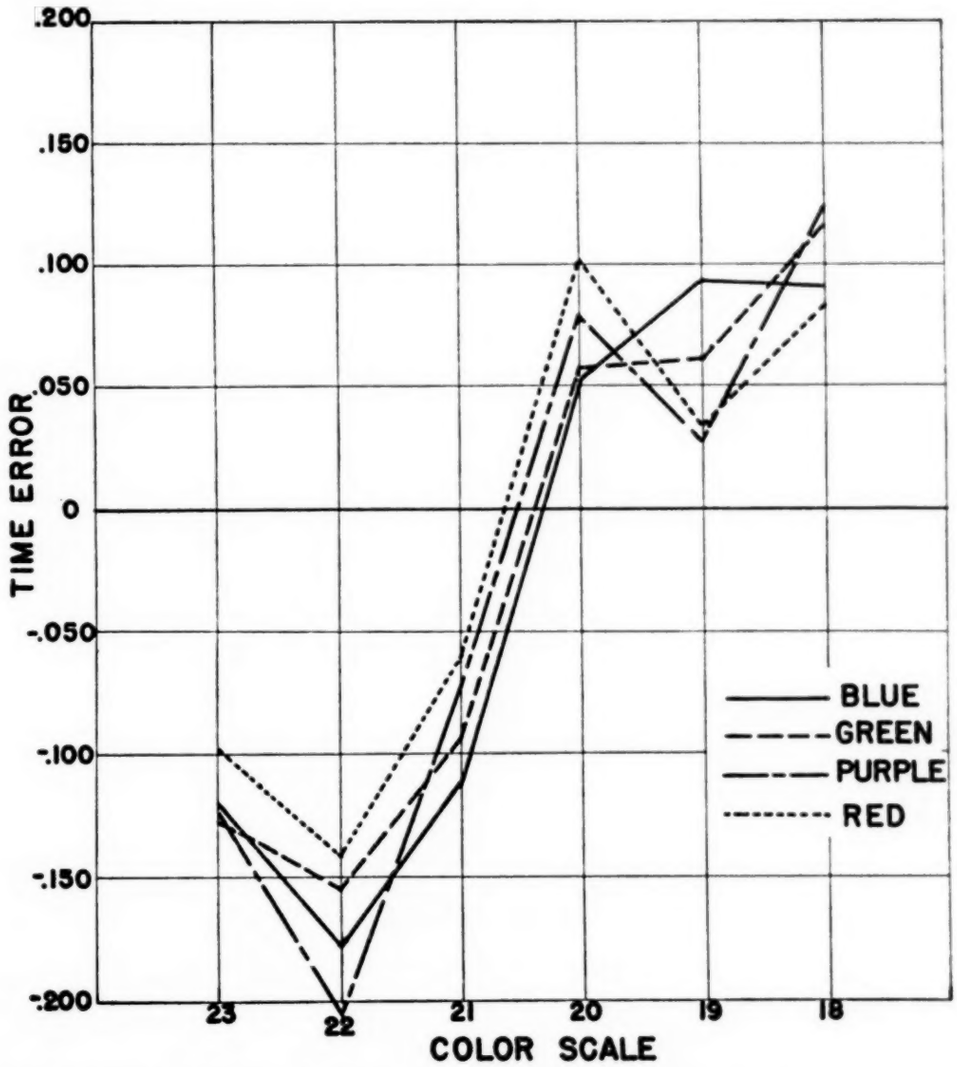


FIG. 1. Variation of time error with serial position for the different colors.

equally difficult; on the other hand they do not seem to affect the time error. Thus, Fig. 1, which graphs the time error for all four colors, each of which occurs in three color combinations, shows no consistent variation. On the other hand there is a definite concomitance between time error and skewness, as may be seen in Fig. 2, in which the ordinate scale for skew-

ness has been multiplied by 5, to permit closer comparison of the two curves. Both curves are negative at stimulus point 23, and pass through a zero somewhere between stimuli 21

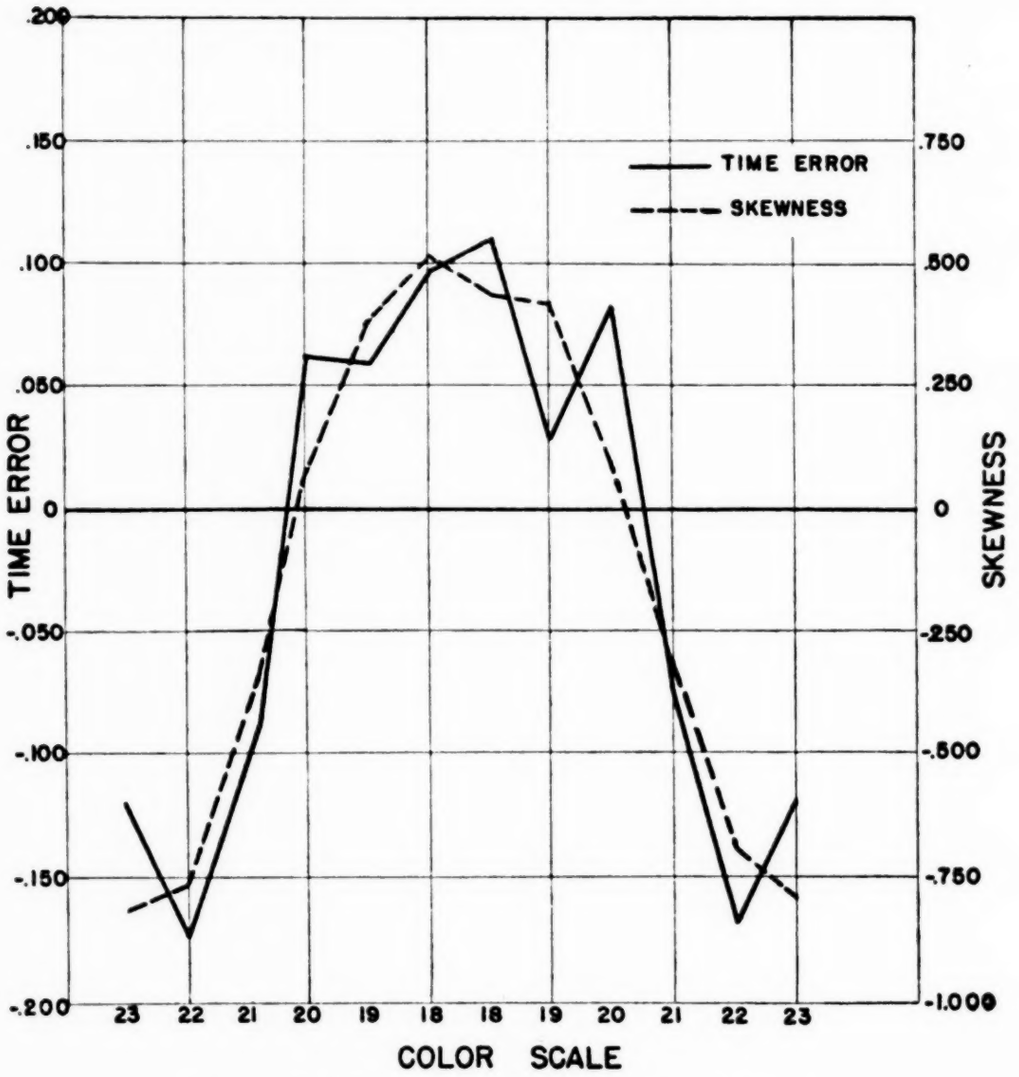


FIG. 2. Relationship between time error and skewness. Skewness scale is five times time error scale.

and 20, beyond which they are positive to the smallest stimuli. From the midpoint of the curves the stimuli are repeated in reverse order to represent the fluctuations of these various factors throughout every point of the scale. It is to be noted that in the ranking method of discriminating color mass, which is in effect a modification of the absolute method, the trend of time error fluctuations is identical with that found by other methods for lifted weights, sound intensities, etc. This study

shows that the skewness of the distributions of judgments consistently follows that trend, through all six color combinations.

Standard Deviations

In Table 4 it may be seen that the standard deviations on the average tend to group rather closely around the value 1.125, with the exception of the standard deviation for stimulus 23. In general also there is an increase of the standard deviation from the terminal value at stimulus 23 to the middle of the series at stimulus 18. The diminution of the standard deviation for stimulus 23 accounts in part for the regression of the time error towards the indifference level at that point, for the standard deviation occurs in the denominator of the fraction from which the time error is computed. It is a point of some interest that there is such a marked difference in time error and in skewness of the distributions while their standard deviations remain closely similar; that is, the variability or spread of the judgments remains fairly constant, though the bias or distortion varies with serial position.

Variation of the Time Error According to Ability of Ss

The Ss were divided into quartiles on the basis of their total error score (Method I), and the distributions of the judgments (Method II) for each quartile were separately compiled. From these distributions, using the D percent method, the time error according to serial position for each quartile is represented in Fig. 3. There is a progressive increase in the absolute magnitude of the time error as one proceeds from Quartile 1 to Quartile 4. The best (most accurate) Ss tend to set their level of reference closer to the standards, which is in conformity with the findings of previous investigators that time errors tend to diminish with practice. Woodrow (7) found that "at every sitting the errors shifted in the positive direction with prolongation of the sitting and at each successive sitting they also became more positive (or less negative) than at the preceding sitting." In this case, with the better quartiles the errors tend to diminish, that is, they gravitate closer to the level of reference. For the best quartiles

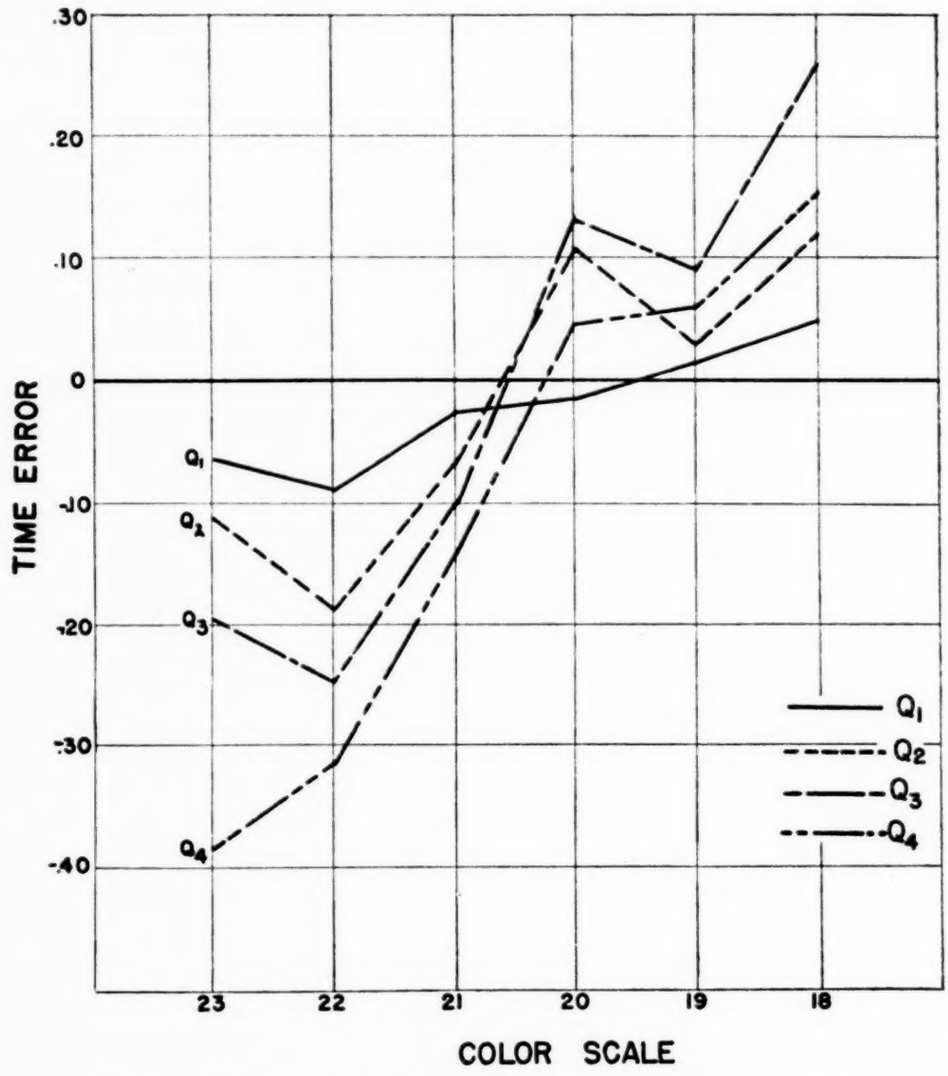


FIG. 3. Time error according to serial position for each quartile. D percent method used to calculate the time error.

the levels of reference are sharper, clearer and closer to the standards; as a consequence they are less subject to distortional effects.

DISCUSSION

The occurrence of the time error in a series of judgments is most intriguing, for it seems to throw light on the nature of the judgment process. Various theories have been advanced to explain this phenomenon but no one theory hitherto presented is satisfactory; though it may adequately explain some phase of the process it does not account for all the phenomena involved. Very probably, on account of the complexity of

the judgment process, the factors to be mentioned below are present in greater or lesser degree in every judgment, and cannot be disregarded.

Some of the explanations advanced are:

(*a*) *Fatigue and Sinking Trace.*—Both theories seek to explain the course of the trace with passing time; the latter theory particularly attempts to account for the trend from negative to positive time errors with an increasing interval between the presentation of the standard and the comparison stimulus. But when all the stimuli are constantly present, as in this procedure, the intervals during which the trace may sink are of the briefest, and the trace may be 'jacked up' as often as necessary, for there is a repeated intercomparison of stimuli. The theory of a sinking trace does not account for the progressive trend from negative to positive in the time error according to serial position. Needham (**8**) has aptly pointed out that the "sinking trace theory implies a constancy and simple determinism which belie the true state of affairs. Indeed even the predisposing influences which are discussed above are relative effects at best; their presence is noted in the appearance of *more* of a certain class of judgments, never in a totality of that class of judgments."

(*b*) *Set* implies the establishment of absolute standards, either prior to the experiment and based upon general experience, or immediately during the course of the experiment. This theory stresses "the level of the set which tends to approximate the average intensity of all the preceding stimuli of the entire series given from the beginning of the experimental sitting, provided the series does not vary over too great a range" (**9**). Absolute standards certainly are set and play a large part in the judgment process, yet Needham's caution is very pertinent here (**10**). "A number of experiments upon the absolute judgment have impressed upon the writer the validity of the suggested 'definition of categories of judgments' or 'construction of an absolute series'; however, the evolution and operation of this subjective scale should be treated in terms of the extreme stimuli and not in terms of a hypothetical standard or mid-point. Stated in simple terms,

the facts seem to be that the O, upon being confronted by a situation in which he must assign certain values to a given series of stimuli appearing in random order, first 'learns' to 'recognize' the boundaries within which he is judging and to assign to these limiting stimuli the relatively more correct judgment." He also quotes Hollingworth's study (11) as an example where the hypothesis of judgment is formulated as a function of the extreme stimuli. Then he points out "It is germane to the trend of the discussion to note that this formulation of the response of the O is in agreement with the O's estimate of his behavior, as well as with the E's evaluation of the O's performance. Thus Ma reported that he 'recognizes the extreme stimuli and guesses at the rest.' . . . (The present discussion) . . . represents an attempt to phrase the theory of judgment more in terms of the O and less in terms of a strict psychophysical relationship. It opens the way for an increased significance of 'set' and 'attitude.' It makes essential a further investigation in which *intraserial effects are emphasized.* (Italics ours.) It does not necessarily exclude the operation of some factor such as that involved in Köhler's trace theory or Lauenstein's gradient theory; it does suggest that such hypotheses must have specific rather than general reference. Finally it raises again the whole question as to the true nature of discrimination with its many related problems."

Obviously the establishment of absolute standards is of first importance, especially for the building up of a frame of reference, a base from which estimations are made. With Needham we agree that this frame of reference, this base, is dependent upon the limiting stimuli rather than upon the average level.

(*c*) *Assimilation and Contrast.*—"Every stimulus that is presented to an observer is brought into relation with the general level of reference, and as a result the judgment will very likely be influenced in some way. . . . Thus the tone is judged in the opposite direction to the level of reference which influences it; this is usually referred to as contrast, while assimilation is the closer approximation of the stimuli that influence each other. . . . Hollingworth attempts to explain

assimilation by the conformity to the law of the central tendency and contrast is the reverse of the phenomenon" (**12**).

Distortions are caused by the presence of stimuli other than those being immediately judged. These extraneous stimuli may belong to the total series of stimuli to be discriminated or may be extraneous, even persisting as background stimuli. One would expect that they would be effective according to their intensity, their deviation from the general frame of reference or the recency with which they have occurred. Studies on retroactive inhibition might by analogy lead us to expect variations in the amount and direction of their effect according to the similarity or dissimilarity of the stimuli with those being judged, the point at which the intruding stimuli are interpolated, and the time of their occurrence. Tendencies have been found for the time error to fluctuate in the direction expected for some of these events. Theories based upon assimilation and contrast are more effective in telling us how distortions of judgment already present may be diminished or accentuated, not how the initial distortions or time errors were originally caused.

(*d*) *End Effect* and *Central Tendency.*—Related to the theories of assimilation and contrast are those based upon End Effect or Central Tendency. They are effective in accounting for the increased magnitude of the time error at the end stimuli of a series, its disappearance (zero time error) at the center of the series, and the migration of the indifference point with the extension of the series in either direction. At best they are theories descriptive of certain phases of the phenomenon —they are not explanatory of its general nature.

(*e*) *Inhibitory Effects by Subsequent Stimuli upon Traces of the Level of Reference.*—There has been an increasing tendency to stress the relationship between memory phenomena and inhibition with the time error. In his summary of time error theories Needham (**3**) points out: "There is the recurrent suggestion that the problem of the time-error has some bearing on the memory phenomena; with the exception of Köhler, recent theorists have tended to avoid the possibility of this consideration. Recent advances in both fields cannot be said

to eliminate this aspect of the problem, and a future establishment of some such relationship seems plausible." In a paper to show that the law of disuse may not be precluded in a study of memory phenomena Pratt (**13**) states the following. "By the use of interpolated stimuli it has indeed been shown that the trace of the standard will move in the direction of the interpolated stimulus. This does not mean, however, that the principle of disintegration must be entirely displaced by that of interaction. Experiments have shown (1) that assimilation to zero (no interpolated stimulus or background) is less than assimilation to any value just above zero, and (2) that assimilation to a trace below the standard is greater than to a trace the same distance above the standard. Neither of these facts is consistent with the principle of interaction. They may easily be accounted for, however, if one assumes that disintegration always operates and that assimilation operates only when some trace is in the comparison-field which is closely enough related to the trace of the standard to permit of interaction. Both principles are needed to explain mnemonic changes: disintegration, always; interaction, whenever certain specified conditions are present." In a recent paper dealing with time order errors in successive judgments and in reflexes Peak (**14**) attempts to relate the two sets of data by consideration of the development of inhibitory effects.

This study shows a close similarity between the functions of skewness and time error in the discrimination of color mass. Obviously skewness is merely a descriptive entity and does not explain the time error; but it does suggest that whatever factors are causing the bias or distortion of the judgments is intimately tied up with the factors causing the time error in the judgments. Moreover these factors must operate progressively and regularly from the extremes of the series to the middle. The introspections of Ss give us a definite clue to the possibility of finding this factor; taken in conjunction with the experimental finding that difficulty in serial discrimination proceeds from the extremes to the center, we are enabled to formulate the following tentative hypothesis for the origin of the time error. A general level of reference, a base or standard

is established at the extremes of the stimuli from which all judgments are gauged and all estimates are made. This general level of reference persists as a trace throughout the whole discriminatory process, though it is subject to continuous modification and readjustment. The traces of the standards by which the stimuli are appraised are sharpest and clearest at the extremes, and possibly that is why the extremes are selected as the reference level. Here the accuracy and precision of the judgments is much greater than anywhere in the series, and the end stimuli are judged with celerity and confidence. Just as the accuracy curves diminish from the extremes to the center, so do the reaction times of the judgments and the degree of confidence with which they are made. To the degree with which the stimuli deviate from the end points, or levels of reference, do judgments based upon them lose in accuracy, clearness, speed and confidence. This lack of clarity and precision of the trace is not due to decreasing intensity, for the middle stimuli are stronger (*i.e.*, the predominance of color mass is greater), than those at one end, nor to recency, for all the stimuli are simultaneously present, but to the presence of other conflicting stimuli which have an inhibitory effect due to their serial position which is more detrimental than the inhibitory effects for the end stimuli.

It is important to recall that the standard of reference is held as a trace, which is in part a memory function, and consequently is subject to fading and inhibition. Striking similarities occur in the curves of recall in serial order and accuracy of judgment in serial order. In the case of memory studies the distortions are recognized as being dependent upon both proactive and retroactive inhibition, the former being due to the inhibitory effect of all stimuli that precede the one being recalled, and the latter being due to all subsequent stimuli in the series.

Analysis of Results

Just how the inhibitory effects of surrounding stimuli exert a biassing influence on the judgment of a particular stimulus, and therefore cause the time error, may be seen by examining these influences for every stimulus in the series. The 23 card

is most correctly judged (81.4 percent of the judgments are correct). The accuracy of the judgment is somewhat beclouded by subsequent stimuli in the series, but as it occurs at the end of the series the trace of the standard is definitely emphasized by the absence of intruding stimuli at one side. Naturally errors can only occur by throwing judgments to the side of the weaker stimuli according to their degree of proximity to the stimulus under question (23 card) which is done in about 18.6 percent of the cases. The 22 card is subject to the disturbing influence of the 23 card on one side, and of all the subsequent weaker ones on the other side, particularly of the 21 card, which is equally far removed from the 22 card as is the 23 card—one serial position. But it is important to note that the trace of the 23 card, since it is closer to the frame of reference, is much sharper than that of the 21 card; hence the 23 card will be less often confused with the 22 card, than will the 21 card—in the general tabulation, Table 1, 435 times as compared with 706, or about 62 percent as frequently. The extra clearness of the 23 card, as compared with the 21 card, when judging the 22 card, is due to the fact that the 21 card is confused in its trace by the greater number of surrounding stimuli. In fact, when judging the 22 card, the 20 card which is two places serially removed, causes more disturbing influence than does the 23 card, which is one place serially removed—455 errors as compared with 435. As a consequence there are far more erroneous judgments to the closed side, in the direction of the center of the series, than there are to the open side, and from the D percent method of calculating the time error the latter will be pronouncedly negative. The added dip in the curve of the time error at the 22 position is partly due to the fact that the trace of the 23 card is so clear that relatively few errors occur—few mistakes are made by throwing the 23 card to the closed side, whereas the 22 card is far more prone to be erroneously judged. The 21 card being more remote from the reference level is less accurately judged than the two preceding. There are two possible erroneous places to the open side, and three to the closed side, with the slight chance (about 1 on 100), that S may not

even judge the card correctly as to color, which adds six more possible places for misjudgment. Even if we compare positions equally removed in serial order from position 21, we see that more mistakes are made by erroneously placing the 21 card in positions 20 and 19, than in positions 22 and 23—(838 and 343 mistakes as compared with 714 and 210 respectively) due to the greater clarity of the traces at the open end and the diminution in inhibiting effect of the surrounding stimuli. As a consequence there will be more errors towards the closed side of the series and the time errors will again be negative, but not so pronouncedly as in the case of the 22 card.

The reversal of the time error occurs with the 20 card. At this point within the same predominating color in the combination there are three places to the open end and only two to the closed end. The trace of the level of reference at the weak end (the closed side), of the stimuli is now coming into force, and the strength of the open end standard is definitely diminished owing to its displacement. If we compare stimuli equally removed in serial position we find that fewer errors are made now on the closed side—for one position displacement, 614 as compared with 808; for two positions removed, 293 as compared with 303 (counting the two 18 positions as the same), and the 106 errors in the 23 position are greater than all the 74 errors on the wrong side of the color combination. Since more mistakes are made on the open side the time error is positive.

The 19 card and the 18 card on the basis of a similar analysis show more errors on the open side than on the closed side, which of course makes the time errors positive. However the magnitude of the positive time error is not as great as that of the negative time error for two reasons: first, there is the possibility of misplacing cards on the wrong color side of the series, greater of course, due to their proximity, for cards on the closed end; and secondly, the trace of the weaker stimuli being closer to the limen of discrimination, the reference level on that side is not nearly so clear as is the reference level on the open side; a comparison of the number of correct judgments for stimuli 23 and 18 brings this out clearly—3903 as

compared with 3021. Since the order of the stimuli is reversed in the second half of the color combination an analysis similar to the foregoing, but in reverse order, would explain both the sign and the magnitude of the time errors.

Summary

The hypothesis advanced to explain the time errors in this experiment postulates the setting up of frames of reference at the terminal stimuli which are inhibited by other stimuli of the series in proportion to their deviation from the frames of reference. It does seem adequate to explain both the sign and the size of the time error according to serial position. Making use of previous hypotheses it indicates the complexity of the judgment process, the presence of levels of reference and sets, and assimilation and contrast effects due to the inhibitory or facilitating influences of the stimuli on each other; moreover it emphasizes the similarity between the judgment process and the memory phenomena, owing to the fading of the trace with time and its tendency to obliteration and distortion by subsequent traces. In the light of this explanation, the central tendency, or point of minimum time error towards the center of the series, is the point usually equally removed from both levels of reference where the inhibitory effects tend to cancel each other. The migration of the point of zero-time error with the extension of the series follows the displacement of the levels of reference. The presence of the gradient or stepwise phenomenon throughout the series is readily accounted for on the basis of this explanation. The analogy between the judgment process and the recall, both of which are influenced by proactive and retroactive effects of similar stimuli, and both of which involve fading impressions with passage of time, suggests explanations of the effects of interpolated stimuli and background stimuli on the time error, and the course of its *p*-function, or fluctuation with time.

In the discrimination of color mass the judgment process is affected by both disintegration and interaction. When the judgment is correct, we are dealing with difficulty according to serial position; when the judgment is erroneous, we are dealing with time errors.

Conclusions

On the basis of some 57,600 judgments of color mass from 32 Ss time errors were determined for serial position.

(*a*) There is a progressive trend from negative to positive time error for serial position, as the discrimination proceeds from stronger to weaker stimuli.

(*b*) This trend is independent of color or color combination, though these vary in difficulty, since the time errors remain closely similar no matter what the color or the color combination in the series.

(*c*) The time error trend is identical in relative magnitude and sign with the skewness trend for serial position.

(*d*) The variability of the distributions of the judgments as determined by the standard deviation varies but slightly with serial position, except for the strongest stimuli. There is some indication of an increasing trend of variability from strongest to weakest stimuli.

(*e*) When Ss are fractionated into quartiles according to error score, the time errors tend to regress towards the indifference level for all serial positions, according to the discriminatory ability of Ss. The regression is greatest for Q_1, the best group, and least for Q_4.

(*f*) An explanation is advanced to account for the time errors based upon levels of reference at the terminal stimuli and distortions in the judgments caused by the inhibitory influences of traces of intervening stimuli.

(Manuscript received March 19, 1940)

References

1. Fernberger, S. W., An introspective analysis of the process of comparing, *Psychol. Monog.*, 1919, **26**, No. 117.
2. Köhler, W., Zur Theorie des Sukzessivvergleichs und der Zeitfehler, *Psychol. Forsch.*, 1923, **4**, 115.
3. Needham, J. G., The time error in comparison judgments, *Psychol. Bull.*, 1934, **31**, 229–244.
4. Philip, B. R., Intra-serial effects in the discrimination of color mass by the ranking method, *J. exp. Psychol.* (in press).
5. Needham, J. G., Contrast effects in judgments of auditory intensities, *J. exp. Psychol.*, 1935, **18**, 215–226.
6. Pratt, C. C., Time errors in the method of single stimuli, *J. exp. Psychol.*, 1933, **16**, 798–815.

7. Woodrow, H., The effect of practice upon time-order errors in the comparison of temporal intervals, *Psychol. Rev.*, 1935, **42**, 127–152.
8. Needham, J. G., Interpolation effects with different time-intervals, *J. exp. Psychol.*, 1935, **18**, 767–773.
9. Woodrow, H., Weight discrimination with a varying standard, *Amer. J. Psychol.*, 1933, **45**, 391–416.
10. Needham, J. G., Rate of presentation in the method of single stimuli, *Amer. J. Psychol.*, 1935, **47**, 275–284.
11. Hollingworth, H. L., The central tendency of judgment, *J. Philos.*, 1910, **7**, 461–469.
12. Long, L., A study of the effect of preceding stimuli upon the judgment of auditory intensities, *Arch. Psychol.*, 1937, No. 209.
13. Pratt, C. C., The law of disuse, *Psychol. Rev.*, 1936, **43**, 83–93.
14. Peak, H., Time order error in successive judgments and in reflexes: I. Inhibition of the judgment and the reflex, *J. exp. Psychol.*, 1939, **25**, 535–566.

STUDIES IN THE TRANSPOSITION OF LEARNING BY CHILDREN: V. THE NUMBER OF STIMULI IN THE TRAINING SERIES AS A FACTOR IN GENERALIZATION

BY T. A. JACKSON AND M. E. ECKHARDT

Columbia University

A theory of transfer in terms of the conditioned response has been set forth by Spence (**6**). In that formulation particular use is made of the concept of generalization. The theory was extended by the senior author (**2**) so as to account for the transition from relative to absolute choice at advanced stages of practice. This was done by adding to Spence's formula a term which made the degree of generalization a function of the amount of practice. That modification of the theory appeared to be adequate to account for the results obtained in strict transposition experiments where it is possible to determine but one point on a generalization curve. It is highly desirable to have some empirical data indicating the exact shape of these curves. This may be obtained by employing a wide range of stimuli in the critical trials rather than the conventional two. If responses were obtained from a number of subjects on a number of critical trials, sufficient data would be had to give an indication of the shape of the whole generalization curve. In the experiment reported here that method was used. The results of an earlier study in the present series (**5**) indicate that transposition experiments can not be accounted for entirely by the conditioning theory unless an additional 'patternedness' factor is employed. In the present experiment another test for the 'P' factor will be made.

A general outline of our experiment follows: Three groups of children were trained to respond to a one-stimulus, a two-stimuli, and a three-stimuli situation respectively. In all cases the stimulus (or stimuli) appeared in one or more of nine squares of a 3 × 3 frame. In Group I (single-stimulus group)

response consisted in selecting the square containing the stimulus. In Group II (two-stimuli group) the stimuli were presented on adjacent squares and the task was to select the square containing the smaller one. In Group III (three-stimuli group) the training was similar and the smallest stimulus was positive. In all three groups the magnitude of the positive stimulus was the same. Following equal amounts of training under each condition, the subjects were required, in the critical trials, to choose among nine stimuli varying about the positive stimulus. By making a frequency distribution of the results an approximation of the shape of the generalization curve may be obtained.

On the basis of the conditioning theory alone, without a patternedness factor, the following type of result is expected. (We do not have sufficient data on the nature of the 'P' factor to predict just how it, if present, might modify these results.) (1) Group I will show a distribution of responses (generalization curve) roughly resembling a normal curve, and it should be centered about the single stimulus used with those subjects. (2) Group II will show a generalization curve similar in shape but displaced toward the smaller stimuli. (3) The distribution of responses for Group III will be still further displaced toward the smaller stimuli. In the latter two groups it is assumed that there is a gradient of excitation about the positive stimulus, and a gradient of inhibition about the negative stimuli. The excitation curve may be assumed to lie in about the same place as does the empirical curve for Group I. The inhibition curve will be centered about the negative stimulus. It, at the maximum, will not be as strong as the positive stimulus at its maximum (6). When these two curves are algebraically summated the resultant is a curve similar in shape to the curve found for Group I, but, as pointed out above, it will be displaced to the left side of the graph which represents the smaller stimuli. In discussing the expected results the terms 'relative' and 'absolute' have not been used. In the present experiment there are so many degrees of relativeness and even of absoluteness that the terms may be used only with caution.

METHOD OF THE INVESTIGATION

In order to secure sufficient data to give significance to an experiment of this type, a substantial number of subjects must be used. The usual method of training individuals separately is so long and tedious that a procedure was devised whereby a group of from 15 to 20 subjects could be trained and given critical trials at one sitting. The method has been described elsewhere (1). It consists in projecting photographic images of the stimuli on a screen; then the subjects designate their choices of the one they think is correct by a sort of voting system; that is, by putting numbered tickets into one of two or more compartments of a ballot box. The details of the materials, apparatus and procedure follow.

Stimulus Materials.—Two sets of stimulus materials were prepared, one for the preliminary trials, and one for the regular training trials and the subsequent critical trials. The stimuli used in the preliminary trials were a cross, a triangle, and an octagon; the cross being positive in all groups. Six other irregular geometrical figures were also used at certain stages of the preliminary trials. These were of the same area as the cross, etc., but were as different as possible in design. In the regular training, square stimuli were used; they were set so that the sides were at a 45° angle with respect to the sides of the background. In the critical trials nine stimuli were employed. Their areas were as follows: No. 1, 66.7 sq mm, No. 2, 100 sq mm, No. 3, 150 sq mm, No. 4, 225 sq mm, No. 5, 337.5 sq mm, No. 6, 506.3 sq mm, No. 7, 759.4 sq mm, No. 8, 1139.1 sq mm, and No. 9, 1708.6 sq mm. Numbers 5, 7 and 9 were used in the regular training, with No. 5 positive in all groups.

All the stimuli were placed in the center of square backgrounds which measured 100 mm on a side. The largest stimulus covered but 17 percent of the background. The relation between the background and the stimulus was probably not an important factor in this experiment; however, this should be tested sooner or later.

The stimulus materials were photographed on 35 mm film and presented to the subjects by means of a projector and screen. For this purpose the stimuli were placed in different arrangements on a frame with nine square holes in it. Each hole was just the size of the backgrounds, 100 mm square, and there were three rows with three squares in each row. The frame itself was 5 mm wide and was made of white cardboard. Only in the stimuli for critical trials were all the squares filled. In the training trials the stimuli were arranged in various combinations according to the following plan: (1) the positive stimulus should appear equally often in each of the nine squares; (2) the positive stimulus should not appear twice in succession in the same square; (3) it should not appear in the various squares in any fixed sequence; (4) when two or three stimuli were presented they should be either horizontally or vertically adjacent; (5) when two or three stimuli were presented the positive should appear an equal number of times in each of the relative positions; that is, to the right, left, upper, lower, or middle position; (6) in the critical trials the positive stimulus should appear at least once in each of the various positions and not more than twice in any position.

Following each stimulus presentation, information was thrown on the screen; the words "THIS WAS THE ONE" appeared in the square in which the positive stimulus had just appeared. Consequently, each alternate frame in the film strip was an information frame. In these frames all the squares but the information square were blank. It goes without saying that no information was given following critical trials.

Response Materials.—The subjects made their choices among the various stimuli by dropping numbered tickets (hat check stubs) into one of nine compartments of a ballot box. The ballot box was made in the same shape as the frame in which the

stimuli appeared, thus making it easy for them to select the slot in the ballot box corresponding to any given square on the screen. All subjects were given tickets numbering from 1 to 100 at the beginning of a session, and all of them were to use ticket number one on trial number one, two on two, etc. Check-ups were made frequently to see that all were using the same numbered ticket. After all trials were over the tickets in the various boxes were collected and checked against the presentation order of the stimuli. In this way it was possible to determine which subjects made errors on which trials and also what choice they made on the critical trials.

The possibility of cheating was eliminated by setting up three-sided screens on each subject's desk. These were made by clipping together two manila folders and setting them on edge with the open side facing the child, the other three sides obscuring his tickets and his ballot box from the view of his neighbors.

Training Procedure.—The children were trained in their own school room. The screen was placed in the front of the room and the projector set at such a distance that the entire screen (30 × 40 in) was just filled by the large square of stimuli. The stimuli, as projected on the screen, varied from 2.25 sq cm to 1463 sq cm; care was taken that each child subject could see the entire screen.

The task required of the child-subjects was explained to them as a type of guessing game. Only nine out of a total of seventy-one failed to learn the task. All children were well motivated. The fun of playing the game seemed to be sufficient to hold their interest for the 40 to 50 minutes of training. Many of them said they enjoyed the game, and none objected to it.

Subjects.—The subjects were children in the third grade of the William Penn Public School and the Hamilton Public School, Mount Vernon, New York. Those who were trained on one stimulus shall be designated as Group I; those trained to select the smaller of two stimuli, Group II; and those trained to select the smallest of three stimuli, Group III. Each one of these major groups was divided into two smaller groups for convenience in testing.

Group I and Group III were from the William Penn School. The groups were equated as nearly as possible for mental age. They were of a low-class socio-economic level. Specific data on the three groups are given in the following table:

TABLE 1

SHOWING NUMBER, SEX, MENTAL AGES AND I.Q. OF THE 3 GROUPS

Group	Total No.	No. Who Learned*	Of Those Who Learned					
			Boys	Girls	Mdn. M.A.	Range M.A.	Mdn. I.Q.	Range I.Q.
I	24	24	11	13	8–8	6.11–10.6	103.5	60–123
II	26	20	7	13	8–3½	7.1 –9.10	95.0	63–119
III	21	18	10	8	8–8	7.1 –10.4	102.5	70–114

* All of the results and discussion will be based only on those children who learned the problem. The criterion for learning was not more than one error in the last 25 trials. One exception, Anthony B., in Group I, made no errors throughout the training except for three errors occurring in the 41–50 trial group.

Results

A summary of the results is shown in Table 2 and in Fig. 1. They are in line with the expectations. The inter-rela-

TABLE 2

Showing the Number and Percent of Responses to Each of the Nine Stimuli in the Critical Trials for Groups I, II and III

Group	Stimuli																	
	No. 1		No. 2		No. 3		No. 4		No. 5		No. 6		No. 7		No. 8		No. 9		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
I	7	2.9	2	.8	0	0	34	14.2	55	22.9	60	25 0	32	13.3	0	0	50	20.8	240	99.9
II	41	20.5	8	4.0	8	4.0	16	8.0	50	25.0	37	18.5	16	8.0	7	3.5	17	8.5	200	100.0
III	74	41.1	5	2.8	11	6.1	40	22.2	9	5.0	11	6.1	11	6.1	5	2.8	14	7.8	180	100.0

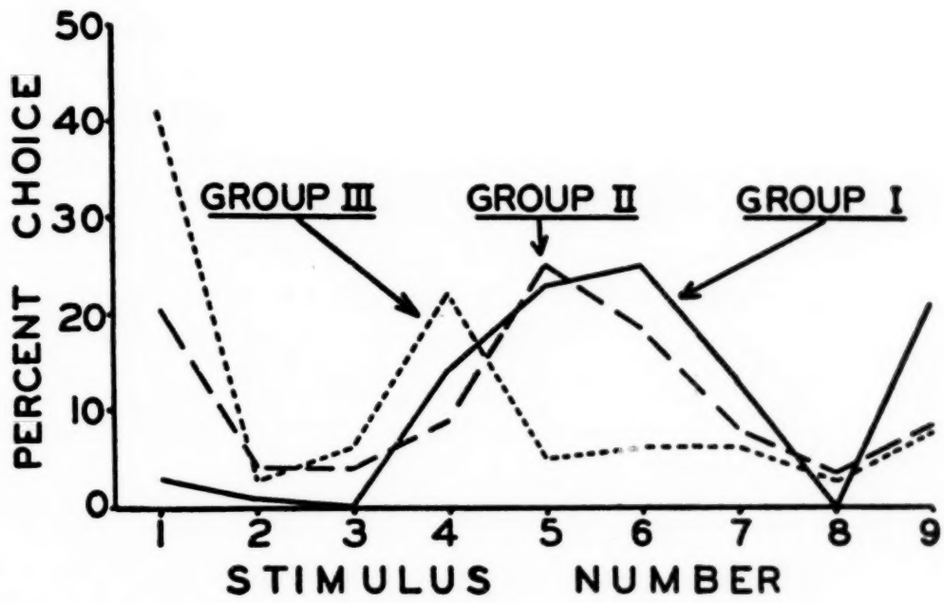


Fig. 1. Showing the distribution of responses ('generalization' curve) for each group.

tionship between the three generalization curves is in general as predicted by the conditioning theory of transposition. However, the up-turns on the ends of the curves were not anticipated, nor was allowance made for the displacement of the curve for Group I toward the larger stimulus, although it is generally known that children have a preference for the larger stimulus. Both of these aspects of the results will be discussed later.

Closer inspection of the data shows that for Group I there were 41.2 percent more choices for stimuli above the positive stimulus than below it. The percent choice above and below were 59.1 and 17.9 respectively. For Group II these percentages were nearly the same, 38.5 and 36.5. Group III showed 49.4 percent more choices for the stimuli below the positive. The differences between these curves may be expressed somewhat more concisely in terms of X^2. The X^2 value obtained when the curves for groups I and II are compared is 73.44; when the curves for groups II and III are compared it is 64.12. Both of these far exceed the value of X^2 for 8 degrees of freedom for a probability of .01. According to this analysis the different sets of data are certainly not due to the same factors.

In the foregoing treatment of data the results for all 10 trials have been thrown together. Although generally the response through a series of critical trials may change due to practice effect within those trials, that did not appear to a significant degree in this experiment. The X^2 test shows that the factors underlying the first 2 trials do not differ significantly from those producing the last 2 trials. The X^2 test for each group yielded values of 2.32, 7.76 and 8.51 respectively for the three groups. These represent probabilities for similar factors of about .96, .48 and .40. This indicates that there were no significant differences between the factors underlying the two sets of data, even if they are not quite the same.

The results have been presented for groups. Frequently significant facts are covered up by such massing of data. Although this experiment has not been caried out with a view to making individual analysis of results, since that has been done elsewhere (**3**), it may be well to include one table which will give an idea of the variety of response within one group. In this table the data have been greatly condensed, and consequently it may require some explanation. For each subject an average response score has been computed. For this purpose response to the smallest stimulus was given a score of 1, and next smallest 2, etc.; the largest stimulus thus would get a score of 9. To show the degree of variability for each subject

TABLE 3

SHOWING THE AVERAGE 'SCORE' AND 'AVERAGE DEVIATION' OF EACH SUBJECT FOR THE TEN CRITICAL TRIALS IN EACH OF THE GROUPS

Subject Number	Group I		Group II		Group III	
	A.M.	A.D.	A.M.	A.D.	A.M.	A.D.
1	1.5	0.7	1.0	0.0	1.0	0.0
2	4.2	0.4	1.0	0.0	1.0	0.0
3	4.3	0.4	1.0	0.0	1.0	0.0
4	4.3	0.4	2.0	1.4	1.0	0.0
5	4.6	0.5	2.4	0.5	1.0	0.0
6	4.7	0.4	4.0	0.2	1.0	0.0
7	4.9	0.2	4.4	0.5	1.9	1.3
8	5.0	0.0	4.6	2.1	2.7	0.4
9	5.4	0.9	5.0	0.0	3.8	0.3
10	5.5	0.5	5.0	0.2	4.0	0.0
11	5.8	0.5	5.1	0.2	4.0	0.0
12	5.9	0.4	5.5	0.5	4.1	0.3
13	5.9	0.2	5.5	3.1	4.7	2.8
14	5.9	0.2	5.6	0.5	5.0	3.8
15	6.0	0.2	5.7	0.4	5.7	1.0
16	6.0	0.0	5.7	0.4	5.7	0.6
17	6.8	0.3	6.5	0.5	6.5	0.6
18	6.9	0.2	7.1	0.5	9.0	0.0
19	7.0	0.0	7.8	0.8		
20	9.0	0.0	9.0	0.0		
21	9.0	0.0				
22	9.0	0.0				
23	9.0	0.0				
24	9.0	0.0				

the A.D. of these same scores was also computed. Table 3 shows these results for each group. An examination of this tabulation makes it clear that the foregoing analysis for the several groups was justified. In no case is the average for a group unduly influenced by the results of one isolated subject.

Since the number of errors made in both the preliminary training and in the regular trials was so small, it is not worth while to compare those data with the type of response in the critical trials.[1]

DISCUSSION OF RESULTS

The essential theory set forth in the introduction has been supported quite well by the data. The fact of the presence of generalization effects among the stimuli employed is clearly demonstrated. Although that theory demands that generali-

[1] The raw data upon which these computations were based may be found in the junior author's M.A. thesis deposited in the Columbia University Library.

zation operate in each individual subject, our results do not specifically show that, because the experiment was not set up with such a purpose in mind. The fact that there is some variation in results between individuals does not necessarily mean that each individual subject did not behave as would be predicted by the theory. Various inhibiting and facilitating factors were doubtlessly at work producing such individual differences among the scores.

Although the three curves representing the results are roughly similar to what may be expected on the basis of the generalization theory, the up-turns at the extremes of the curves are not accounted for by that formulation. Since this phenomenon may be of more than minor importance, it will be given some consideration in this discussion. It does not seem likely that the distribution of results is due merely to a ceiling or floor effect. Rather this trimodality suggests the presence of more than one type of factor operating in producing these results. The component involved, in addition to the generalization mechanism, may be the 'patternedness' factor which was mentioned in an earlier report (**5**). If such is the case it would operate somewhat as follows for an individual subject. He quickly learns to select the *smaller* stimulus in the training situation, but when the critical trials are presented 'smaller,' as applied to nine stimuli, becomes 'smallest.' In fact, what he learned in the training situation was not to select a specific stimulus rather than this other one, but *smaller* in the broadest, vaguest and most general sense. What the subject learns is little more than an orientation; as applied to these stimuli it is a directional tendency; that is, a tendency toward the smaller (smallest) stimuli. This is supported by the observation that stimuli may be transposed after very slight training and the subject will make a relative choice without realizing that any change was made in the stimuli. After more training the relative response will be made as it has been observed in the transposition experiments. If training is continued still further, the absolute choice will be made (**2**, **3**, **4**).

What seems to occur is something as follows: At first only

a vague pattern is learned—a directional tendency; if critical trials are given at this stage of practice, the subjects will make a 'directional' relative choice—the limiting case. As training is continued, this type of choice wanes and another takes its place. The latter, presumably, is produced by the conditioning mechanism set forth in Spence's theory, and it is this one that accounts for the 'transpositional' relative choices, as well as the absolute choices. Certainly there is no patterning factor operating when the choice has ascended to the absolute stage; but at just what stage of practice its effect was lost is not clear.

We wish to tender the foregoing interpretation as very tentative, and hope that future experimentation may clarify the exact relation that exists between the two mechanisms.

Summary and Conclusions

Three groups of third grade children (about 20 in each group) were trained, by the group method, to make a simple size discrimination between visually presented stimuli. Group I was trained to react to a given stimulus (stimulus No. 5 of an array of 9), which was presented in one of the squares of a frame containing nine stimulus 'grounds'; the eight other square grounds were blank. Group II selected the same stimulus (No. 5) when it was paired with a larger one (No. 7). Group III selected the positive stimulus when it was the smallest of a trio (Nos. 5, 7 and 9). Following 65 trials of practice all subjects were given 10 critical trials, the whole thing being carried out in one session. In the critical trials 9 stimuli were presented (Nos. 1–9) simultaneously. In this fashion it was thought generalization could be studied for all three types of training situations.

The following conclusions are drawn from the data:

1. A wide range of responses were made by all groups, thus indicating a generalization effect.

2. The curve for Group I was not centered upon stimulus No. 5, but was displaced to the right (toward larger stimuli), thus indicating an initial preference for larger stimuli.

3. The curve for Group II was similar in shape, but moved

to the left of curve 1, thus showing the effect of the negative stimulus in the training.

4. The distribution curve for Group III is still further displaced to the left, indicating a summation of inhibition effects generated by the two negative stimuli in the training.

5. The up-turns at the extreme ends of the curves are interpreted as being due to a 'patternedness' factor, the influence of which is probably exerted before the generalization mechanism operates to a significant degree.

6. It may be inferred that in the absence of a dip in the two latter curves just above the negative stimuli, the inhibition effects induced by the negative stimuli must have been spread over a range of adjacent stimuli. The resulting distribution curves are accounted for by the algebraic summation of generalized excitation effects (generated by the positive stimulus) and generalized inhibition effects (generated by the negative stimuli), plus the effect of the 'patternedness' factor.

(Manuscript received April 19, 1940)

References

1. Jackson, T. A., A photographic method for studying discrimination-learning in children, *J. exp. Psychol.*, 1940, **26**, 116–123.
2. Jackson, T. A., Stonex, E., Lane, E., and Dominguez, K., Studies in the transposition of learning by children: I. Relative vs. absolute response as a function of the amount of training, *J. exp. Psychol.*, 1938, **23**, 578–600.
3. Jackson, T. A., and Dominguez, K., Studies in the transposition of learning by children: II. Relative vs. absolute choice with multi-dimensional stimuli, *J. exp. Psychol.*, 1939, **24**, 630–639.
4. Jackson, T. A., Studies in the transposition of learning by children: III. Transpositional response as a function of the number of transposed dimensions, *J. exp. Psychol.*, 1939, **25**, 116–124.
5. Jackson, T. A., and Jerome, E., Studies in the transposition of learning by children: IV. A preliminary study of patternedness in discrimination learning, *J. exp. Psychol.*, 1940, **26**, 432–439.
6. Spence, K. W., The differential response in animals to stimuli varying within a single dimension, *Psychol. Rev.*, 1937, **44**, 430–444.

THE RANGE OF APPREHENSION AND SENSORY DISCRIMINATION

BY LEO M. HURVICH

Harvard University

Hunter and Sigler (7) have reported the results of an experiment on the range of apprehension.[1] Eschewing the traditional approach, they have considered the problem as one on visually controlled behavior. More interesting still, they have treated it as a problem in visual discriminative behavior to be explained in terms of specific visual mechanisms, and have determined the extent to which the Bunsen-Roscoe Law is applicable to such data. The range of apprehension is here defined as the number of dots identified correctly 50 percent of the time by an observer for stated stimulus conditions. For purposes of analysis, they considered the limen thus defined to be a constant effect and determined the relationship existing between various durations and intensities of the stimulus when both of these physical variables were systematically varied throughout a considerable range.

The experimental findings of Hunter and Sigler indicate that within certain limits, the Bunsen-Roscoe Law ($I \cdot t = C$) is applicable to the range of apprehension. Since inspection of their graphs for spans of from 1 to 7 dots reveals a striking similarity to graphs for data on visual discrimination, considerations of parsimony lead these authors to apply to their own data the photochemical explanations commonly advanced to explain the visual discriminatory functions.

The writer is in complete accord with Hunter and Sigler's thesis that the phenomena of range of apprehension should be related to the known visual mechanisms and interpreted in terms of them. The following experiment tends to confirm

[1] We are employing the conventional phrase rather than Hunter and Sigler's suggested 'span of visual discrimination' simply to avoid confusing the latter with 'visual discrimination.'

Hunter and Sigler's general view, but also raises a question as to the necessity of their specific interpretations.

Since visual sensitivity is greater for binocular than for monocular perception (**9**, **10**),[2] then, if the range of apprehension is a function of visual discrimination, it should follow that the range for binocular is greater than for monocular vision. To test the validity of this assumption the range of visual apprehension was determined for both binocular and monocular stimulation with conditions otherwise constant.

The stimulus cards (5–10 dots)[3] were presented by means of a modified Dodge tachistoscope at a fixed illumination (1.2 f.c.) and for a time interval fixed at 100 ms.[4] The stimulus field could be presented to either eye alone by occluding the proper opening in the stereoscopic hood (without lenses), or to both eyes, by removal of the masking card. The usual experimental variations and precautions were introduced and observed within any given session, and to compensate for practice effects from session to session.

Determinations were made for six observers (**14**). In general the total number of correct judgments was found to decrease as the number of dots increased in both monocular and binocular observation. A comparison of the limens obtaining under these two conditions, however, indicates that the results are not univocal. Each stimulus was presented 18 times and the results when graphed are not so smooth as those of Hunter and Sigler (35 trials).

These limens were calculated from frequency distributions. Reaction times, however, also measure sensitivity (**13**; **15**, 333 ff., 426 ff.). Since the subject's responses were in general shorter and given with greater subjective certainty for binocular stimulation than for monocular stimulation, it was decided to time them. The times were measured by a voice-key and a chronoscope wired in series with the tachistoscope. The observer was instructed to react 'as quickly as is convenient rather than as quickly as possible.'

Times for two observers are presented in Table I, which shows (1) that the average reaction-time for a correct response varies directly with the number of dots on the card (*cf.* **13**), (2) that the values for each eye measured independently are not the same, and (3) that the average reaction-time for a correct response tends to be shorter for binocular than for monocular vision (19 instances out of 22). Although the total number of cases is small, and the averages are necessarily based on an unequal number of cases, the general trend is

[2] Since the completion of the experiments reported here, Crozier and Holway have confirmed these findings (**5**, **6**). However, *cf.* (1).

[3] The dots were .2 cm in diameter and viewed at a distance of 46 cm (visual angle = 15′). They were scattered over an area subtending 2° of arc.

[4] A more complete description of the apparatus can be found in (**3**, 167–168).

TABLE I

AVERAGE REACTION-TIMES FOR CORRECT RESPONSES
B = binocular, L = left eye, R = right eye. Reaction time in millisec.

No. of Dots	D.N.J.			A.H.B.		
	B	L	R	B	L	R
5	2.08	2.14	2.16	2.47	2.57	2.56
6	2.52	2.71	2.57	2.74	2.96	3.06
7	2.83	3.37	2.85	3.52	3.28	3.27
8	3.37	3.38	3.74	3.43	4.18	3.92
9	4.06	4.64	4.03	3.62	3.95	4.15
10	4.97	5.55	5.46	—	—	—

nevertheless clear. This result is in conformity with expectation—an expectation based on the assumption (*vide supra*) that the process is basically one of sensory discrimination.

This same result, however, calls into question the necessity of Hunter and Sigler's further specification of the mechanism as photochemical in nature. They say: "The span of discrimination of from 1–7 dots, which is a single discriminatory event, is determined by the photochemical processes in the eye up to that duration at which the sensory contribution is complete," and they add that, the 'span of attention' is "a form of behavior which in the field of vision is controlled on the receptor side by time and intensity factors" (**7**).

In contrast to this emphasis on peripheral events it has been pointed out that the fact that visual differential sensitivity is greater for binocular than for monocular vision "creates a strong presumption that the physiological processes critically responsible for the discrimination of a just noticeable difference in brilliance are located in the central nervous system . . ." (**8**). The results presented here are consistent with this belief; the events crucial to the determination of the single discriminatory event may be localized in the central nervous system.[5]

[5] Other factors which are less well understood seem also to be operative in addition to the simple sensory mechanisms. Hunter and Sigler indicate that for spans of 8 dots or more the situation is more complicated in that counting enters the picture and more than one event is being dealt with at a time. Other evidence of this greater complexity is seen in the differences Warren obtained between simple cognitive reaction-times and counting responses for the same stimulus (**13**). Furthermore, a consideration of the

The significance of this conclusion lies not only in its application to the specific functions investigated by Hunter and Sigler; it has wider implications, in that it throws doubt on the necessity of the photochemical interpretation as applied to visual sensory functions in general.[6] Hunter and Sigler's argument is essentially this: the functions for the range of apprehension are (within limits) similar to the visual discriminative functions; the forms of the latter are completely determined by photochemical events; therefore, the same interpretations are applicable to the phenomena of the range of apprehension. This conclusion follows if, and if only, the second premise is true. If, on the other hand, Hunter and Sigler's functions for the range of apprehension are primarily centrally determined, a presumption favored by the present results, then it follows that the functions for sensory discrimination (in the intact human organism) are primarily centrally rather than peripherally determined also.[7] Thus, it is possible completely to reverse their argument from similarity.

Summary.—Hunter and Sigler have shown that the Bunsen-Roscoe Law is applicable (within limits) to the data on the span of visual apprehension. From this they conclude that the span of apprehension resolves itself into a span of visual discrimination and is thus to be explained in terms of photochemical retinal processes. On the basis of experimental evidence which indicates that the span of apprehension is sig-

fact that reaction-time tends to increase (decreased sensitivity) as the total number of dots on a card increases (*cf.* Table I), while differential sensitivity increases with an increase in area (**6, 8**) leads to similar conclusions.

The experiment here reported tends to support Boring's view that attention is a problem in organization. The fact that the binocular range is 'greater' than the monocular range (for this is the sense of our results) fits nicely the view that "Attention is limited in range by the amount of [brain] tissue available for functional differentiation in the perceptual or ideational whole" (**2**, 198).

[6] Nor does such a statement imply that we are not dealing with a complete 'response arc' in the intact organism (*cf.* **11**, 156–160) for, as Troland has pointed out, "Our primary problem is, of course, to establish the psychophysiological relationships which exist between visual experience and each of [the] successive stages in the afferent arc mechanism" (**12**, p. 52). For a theoretical development of the problem of central vs. peripheral determination *cf.* (**11**, ch. 6).

[7] *Cf.* Crozier's analysis of the data used to test the validity of the Bunsen-Roscoe Law for photic stimulation. The functions' being fitted by a normal probability integral leads to a discounting of the usual photochemical interpretations (**4**).

nificantly different for monocular and binocular vision, it is argued in the present paper that although the processes are sensory, the critical determinants of the span may be central in nature and not peripheral at all. The validity of the photochemical equations as quantitative explanations of visual discriminative functions in general is also to be scrutinized in the light of this analysis.

(Manuscript received March 26, 1940)

References

1. Bartlett, N. R., and Gagné, R. M., On binocular summation at the threshold, *J. exp. Psychol.*, 1939, **25**, 91–99.
2. Boring, E. G., *The Physical Dimensions of Consciousness*, New York, 1933.
3. Chapman, D. W., Relative effects of determinate and indeterminate *Aufgaben*, *Amer. J. Psychol.*, 1932, **44**, 163–174.
4. Crozier, W. J., The theory of the visual threshold. I. Time and intensity, *Proc. nat. Acad. Sci.*, 1940, **26**, 54–60.
5. Crozier, W. J., and Holway, A. H., Theory and measurement of visual mechanisms. II. Threshold stimulus intensity and retinal position, *J. gen. Physiol.*, 1938–1939, **22**, 351–364.
6. Crozier, W. J., and Holway, A. H., Theory and measurement of visual mechanisms. III. ΔI as a function of area, intensity and wave-length, for monocular and binocular stimulation, *J. gen. Physiol.*, 1939–1940, **23**, 101–141.
7. Hunter, W. S., and Sigler, M., The span of visual discrimination as a function of time and intensity of stimulation, *J. exp. Psychol.*, 1940, **26**, 160–179.
8. Holway, A. H., and Hurvich, L. M., Visual differential sensitivity and retinal area, *Amer. J. Psychol.*, 1938, **51**, 687–695.
9. Lowry, E. M., Some experiments with binocular and monocular vision, *J. opt. Soc. Amer.*, 1929, **18**, 29–40.
10. Lythgoe, R. J., and Phillips, L. R., Binocular summation during dark adaptation, *J. Physiol.*, 1938, **91**, 427–436.
11. Troland, L. T., *The Principles of Psychophysiology*, v. I, New York, 1929.
12. Troland, L. T., *The Principles of Psychophysiology*, v. II, New York, 1930.
13. Warren, H. C., The reaction-time of counting, *Psych. Rev.*, 1897, **4**, 569–591.
14. Wolfson, W. Q., *On the range of visual apprehension: I. Monocular vs. binocular observation.* Honors thesis. Harvard College, Department of Psychology, April 1938.
15. Woodworth, R. S., *Experimental Psychology*, New York, 1938.

THE RELATION BETWEEN REACTION TIME AND TEMPORAL LOCATION OF THE STIMULUS ON THE TREMOR CYCLE

BY JOSEPH TIFFIN AND F. L. WESTHAFER

Purdue University

It is a commonplace observation that reaction time varies not only from one subject to another but also from one trial to another for a given subject during a single testing period. These variations take place in spite of what is assumed to be adequate control of the experimental situation, and in spite of introspective assurance from the subject that he is as 'ready' or 'set' one time as another. When such variations in the result of an experiment take place, it is a safe assumption that *something* is changing from one trial to another—that the conditions are not as constant as we have supposed them to be.

The present experiment was formulated from the hypothesis that one uncontrolled variable which may have an effect on the reaction time is the temporal location of the stimulus on the tremor cycle of the reacting muscle group. Travis (1) has shown that there is a relation between voluntary movement and tremors. He reports that "in the great majority of instances a sudden voluntary movement blends into the tremor rhythm so as to be a continuation of the downward or upward phase of the tremor."

The purpose of the present investigation was to determine:

(1) Whether average reaction time differs significantly with different temporal locations of the stimulus on the tremor cycle, and

(2) Whether certain temporal locations of the stimulus on the tremor cycle result in less variation in reaction time for a given subject from one trial to another than do other (or random) locations.

If a location can be found which results in significantly shorter reaction times than when this element is uncontrolled,

it is reasonable to expect that reaction times obtained when the stimulus is kept at this location should show less variability from one trial to another than reaction times obtained when this factor is left uncontrolled. This would follow because with the elimination of this uncontrolled variable which affects the results, the reaction time should be affected by one less uncontrolled variable.

Apparatus

The apparatus was set up in such a fashion that the stimulus could be initiated (1) on the top phase of the tremor cycle, (2) at random on the tremor cycle, and (3) on the bottom phase of the tremor cycle.

The apparatus consists of two standard tambours, each clamped to a ringstand, connected by $\frac{1}{4}''$ rubber tubing 18″ in length. Very thin rubber sheeting was used for the tambour diaphragms in order to make the moving parts of the system as light as possible. To this end, also, the tambour lugs were made of stiff cardboard, for bases, upon which were glued short ($\frac{1}{4}''$) lengths of wooden lugs in an upright position. Utilizing the regular tambour saddle, the wire pointer was removed and a broomstraw 4″ in length was fitted in its place. A copper wire of very small diameter was then wound about the length of the straw and extended, at the end of the straw, to project out $\frac{1}{4}''$. The other end of the wire was grounded against the tambour saddle.

A screw turning device was clamped to another ringstand and two heavy copper wires 4″ in length were fastened to the device in a V-shaped fashion. The ends of the V were allowed to project in such a manner that one wire was above and the other below the $\frac{1}{4}''$ projection of wire on the broomstraw. A 'contact' key was placed at the base of this latter ringstand.

Behind the first tambour was placed a 'break' key in such a manner that its lever extended over the tambour drum, almost touching the tambour lug.

A circuit was then set up (see wiring diagram, Fig. 1) in which a $22\frac{1}{2}$ volt battery was the source of current. In this circuit a 6–8 volt light bulb (in socket), a chronoscope which records the time in hundred and twentieths of a second, and a small double contact relay were placed in such a manner that when contact was made against the wire on the broomstraw by either of the two wires in V-shaped position, or when contact was made by the 'contact' key, the light would flash on (remaining on because of the action of the relay) and, simultaneously, the chronoscope would begin to record. The 'break' key behind the first tambour was placed in the circuit in such a manner that it would break the entire circuit, shutting off both the light and the chronoscope. A wallboard shield was placed in the apparatus between the position of the subject and the operator.

When the subject places a finger upon the lug of the first tambour, the finger tremors of the subject set up air currents in the rubber tubing which in turn make an amplified reproduction of the tremors on the broomstraw pointer mounted on the second tambour. These fluctuations are in opposite phase to the original tremor (due to the mechanics of the air tube and tambours).

By means of this double tambour apparatus, the stimulus may be given on the top phase of the tremor, bottom phase of the tremor, or at random with respect to the tremor cycle. In giving the stimulus on the top phase of the tremor, the screw turning

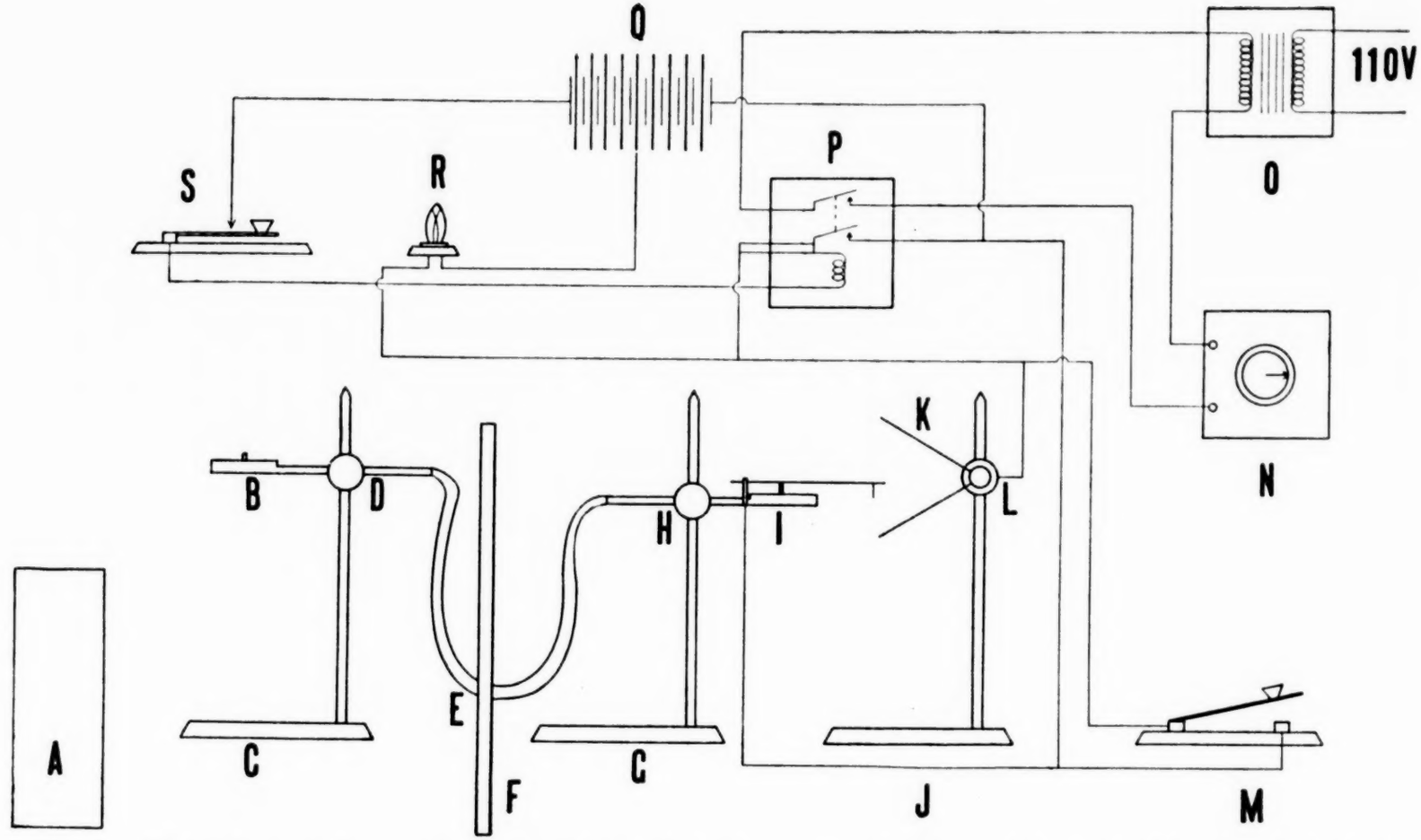


FIG. 1. Wiring diagram of apparatus for determining relation between reaction time and the temporal location of stimulus on tremor cycle. Showing: *A*, Arm Rest; *B*, *I*, Tambours; *C*, *G*, *J*, Ringstands; *D*, *H*, *L*, Ringstand Clamps; *E*, Rubber Tubing; *F*, Wallboard Shield; *K*, V'd Wires Contact on Turning Device; *M*, Contact Key; *N*, Chronoscope (Cenco Impulse Counter); *O*, 110–18 v. Transformer; *P*, Double-Contact Relay; *Q*, 22½ v. Battery; *R*, 6–8 v. Light; *S*, Breaker Key.

device, indicated by *K* in Fig. 1, is rotated until the bottom of the two V-shaped wires touches the arm of the tambour. When this contact is made, the light and the chronoscope are started simultaneously. Since this contact must be made at the bottom phase of the excursion of the tambour arm (which will correspond to the upward excursion of the tambour *B* whose lug is in contact with the finger), the stimulus will necessarily be given on the upward or top phase of the tremor cycle. In like manner, if the screw turning device *K* is rotated in the opposite direction, the stimulus will be given on the opposite or bottom phase of the tremor cycle. In giving the stimulus under the two above mentioned conditions, the screw turning device *K* is rotated by the experimenter so as to make contact only when definite tremors are in evidence. The presence of the tremor may readily be observed by watching the lever of tambour *I*. For the stimuli given at random on the tremor cycle instead of using the screw turning device *K* the contact key *M* is used. Since this contact is in no way related to the tremor cycle, it is obvious that a series of contacts made through this key will be unrelated to any movements of the finger.

In order to keep constant the conditions of stimulation from series to series, the experimenter watched the lever of tambour *I* and gave no stimuli through the contact key *M* unless tremors were in evidence at the instant of contact. During the experiment, the subject is unaware of the mode of initiation of stimulus. He simply observes the light, and, as described later, reacts as quickly as possible without knowing that in two thirds of the trials he himself is initiating the stimulus through involuntary movements of the finger.

Upon seeing the light flash on, the subject quickly depresses the finger on the first tambour. As the finger strikes the 'break' key, the circuit is broken, shutting off both light and chronoscope.

The length of time the light and chronoscope are on is the subject's reaction time for that trial.

Refinements of the Apparatus

Since it was desired that the fluctuations of the broomstraw pointer be exactly one half cycle opposite the position of the finger in its tremor cycle, it was necessary to determine the optimum length of tube which should join the two tambours to produce this result.

With special pointers in the tambour saddles, the tambours were set up one over the other against a kymograph drum and records made using various lengths of tubes. It was found that up to three times the length of the first tube (18″) there were no observable differences in the phases of the tremor curves. That is, the peak of the one cycle record fell exactly vertically below the very bottom of the trough of the other cycle record. At four times the length of the tube, a slight lag in the second tambour was noted. From this preliminary experimentation it was concluded that a tube 18″ in length was sufficiently short so that the time lag between the two tambours could be disregarded for the purposes of the present experiment.

It was also necessary that the light stimulus occur at the instant of contact. To discover if this was the case, the stimulus light was flashed in front of (and focused upon) an aperture in front of a moving photographic film. A hole was cut in this film and a copper wire, serving as a brush, made contact with a copper plate through the film when this hole reached the aperture. A 100 cycle time line was simultaneously photographed on the film. The developed film thus revealed the lag of the stimulus light after contact is made. It was found that this lag could be reduced to .005 seconds by using 9 volts of current on the 6–8 volt bulb. Nine volts were therefore used to light this bulb throughout the experiment.

In order to determine the amount of lag in the circuit caused by the relay, a circuit was set up in which the relay was placed between an electric clock and the chronoscope. A contact key and a 22½ volt battery were included in the circuit. Thus, the difference in readings between chronoscope and the electric clock yielded the amount of lag caused by the relay. Thirty readings were made and the differences averaged. The average amount of lag due to the relay was found to be .055 seconds. This amount was added as a constant to the readings obtained during the experiment.

Procedure

The subjects for this study were selected from graduate and undergraduate students by their willingness to engage in the experiment. Each subject was seated comfortably before the apparatus. The operator was seated to the right of the subject at the contact controls of the apparatus. His actions were screened from the subject by the wallboard shield.

The subject was instructed to place his right arm on the arm rest provided, and extend his right hand with the middle finger over the tambour lug. This was demonstrated, if necessary. The subject was then given these instructions: "Bring your middle finger down slowly until it *lightly* touches the lug. Keep your finger *lightly* on the lug until the light flashes on—then, as quickly as you can, press your finger down! When your finger presses the long lever down, the light is shut off. You must press down as quickly as you can after seeing the light because the length of time the light is on is the length of your reaction time for that trial."

The subject was then told to listen for the verbal 'Ready' signal before each trial, in order that he be prepared. He was also told that the finger need not be poised over the lug between trials, but could be relaxed until the next 'Ready' signal.

Several trials were made in order to determine if the subject was following instructions correctly. These trials were not recorded.

At the 'Ready' signal, the subject began to produce tremor fluctuations on the straw pointer. The operator then moved the lower V'd wire up until contact was made against the broomstraw pointer at the lowest point of fluctuation.

This contact activated the light and chronoscope simultaneously. The subject then pressed down the lever shutting off the light and chronoscope. The operator read the chronoscope and recorded the time in the proper column on a score sheet. The same procedure was followed for all trials; the only difference being the type of contact made.

For all subjects, the contacts were made in the same order which was: first, top phase of tremor; second, random phase of tremor; third, bottom phase of tremor. This order was repeated until the series was completed.

Fifty subjects were used in the experiment. Each subject was given 20 trials for each stimulus location, or 60 trials in all.

Results

The results for mean reaction time and variability of reaction time under the different methods of stimulation are given in Table I.

The differences between the figures shown in Table I cannot be evaluated without considering the correlations between reaction times under the several conditions studied.

TABLE I

MEAN REACTION TIME OF THE 50 SUBJECTS WHEN THE STIMULUS IS LOCATED AT DIFFERENT POINTS ON THE TREMOR CYCLE

	Stimulus at Top Phase on Tremor Cycle	Stimulus at Bottom Phase of Tremor Cycle	Stimulus at Random on Tremor Cycle
Mean	.193	.193	.203
σ	.028	.029	.030

These correlations are shown in Table II.

TABLE II

INTERCORRELATIONS OF MEAN REACTION TIMES WHEN THE STIMULUS IS LOCATED AT DIFFERENT POINTS ON THE TREMOR CYCLE

	Stimulus at Bottom Phase of Tremor Cycle	Stimulus at Random Phase of Tremor Cycle
Stimulus at top phase of tremor cycle	$.95 \pm .015$	$.94 \pm .018$
Stimulus at bottom phase of tremor cycle		$.96 \pm .012$

From the data tabulated in Tables I and II it was possible to determine the significance of differences in mean reaction time and variability of reaction time under the different conditions of stimulation. These calculations are given in Table III.

TABLE III

DIFFERENCES BETWEEN MEANS AND S.D.'S OF REACTION TIMES OBTAINED UNDER DIFFERENT TEMPORAL RELATIONS BETWEEN STIMULUS AND TREMOR CYCLE

	Difference in seconds	$\sigma_{diff.}$	Critical Ratio
Reaction times obtained when stimulus is given at random are, on the average, longer than when stimulus is given at the top of the tremor cycle by	.010	.001	10
Reaction times obtained when stimulus is given at random are, on the average, longer than when stimulus is given at the bottom of the tremor cycle by	.010	.001	10
Reaction times obtained when stimulus is at the top of tremor cycle are, on the average, the same as those obtained when stimulus is at the bottom of the tremor cycle	.000		
S.D. of reaction times when stimulus is given at random is greater than when stimulus is given at the top of the tremor cycle by	.002	.0010	2
S.D. of reaction times when stimulus is given at random is greater than when stimulus is given at the bottom of the tremor cycle by	.001	.0012	.8
S.D. of reaction times when stimulus is given at the bottom of the tremor cycle is greater than when stimulus is given at the top of the tremor cycle by	.001	.0013	.8

The figures given in this table show that average reaction time is shorter by a slight (but significant) amount when the stimulus is given at the top or bottom of the tremor cycle than when it is given at random. The data also show that the variability under the two first named conditions is slightly less than that under the latter condition, though the differences in variability are not statistically significant. The tabulated data also show that there is no difference between average reaction time for the top phase of tremor stimulus and the bottom phase of tremor stimulus. There is a difference, but not a significant one, between the variability of the former and latter reaction times, the reaction times from the bottom phase of the tremor stimuli being slightly more variable.

Summary

This study has demonstrated that for a group of 50 subjects, the mean reaction time is significantly shorter when the stimulus is given on the top or bottom phase of the tremor cycle than when given at random with respect to the tremor cycle. In other words, one heretofore uncontrolled factor which accounts for a part of the variation in reaction time from trial to trial is the temporal location of the stimulus on the tremor cycle. As to be expected when a new (and influencing) factor is placed under control, the variability in reaction time is smaller when the stimulus is given at either extreme of the tremor cycle than when given at random. The reduction in variability, though slight, is such that there are 98 chances in 100 that it is not due to chance when the stimulus is given at the top of the tremor cycle, and 79 chances in 100 that it is not due to chance when the stimulus is given at the bottom of the cycle.

The reduction in variability, though slight and not conclusively significant statistically, is in the direction we would expect when a determining factor which has previously been allowed to vary randomly is placed under control.

(Manuscript received April 4, 1940)

Reference

1. Travis, L. E., The relation of voluntary movement to tremors, *J. exp. Psychol.*, 1929, 12, 515–524.

EFFECT OF BENZEDRINE SULPHATE ON BLOCKING IN COLOR NAMING *

BY RALPH F. BERDIE

University of Minnesota

Benzedrine is a recently discovered compound of the epinephrine series. Its chemical formula is $C_6H_5CH_2CHNH_2CH_3$ and its chemical name is phenylisoprophylamine. Preliminary experiments have suggested that the drug is a stimulant which, in most instances, has apparently failed to leave any after-effects.

Its physiological effects have been reported by Guttmann and Sargent (**6**) and Peoples and Guttmann (**13**). Twenty m.g. or more raises the blood pressure appreciably, with the period of latency between 45 minutes and two hours and the maximum effect appearing about one hour later. When taken on an empty stomach, the speed of absorption tends to produce dizziness, shivering, palpitation, tremor and anorexia. It also tends to inhibit urination and evacuation of the bowels (**12**).

Reports of the psychological effects of the drug are somewhat contradictory. Peoples and Guttmann found it produced an increase in the number of continuous additions. McNamara and Miller (**8**), using smaller dosages, discovered no effect upon three digit multiplication. Molitch and Eccles (**9**) found the drug produced higher scores on retests with intelligence tests and Molitch and Sullivan (**10**) noted that it produced higher scores on retests with the new Stanford achievement tests.

Gwynn and Yates (**7**) give the reactions reported by 14 college students given doses of ten m.g. benzedrine sulphate. They reported temporary increase in energy, increased power

* This study was done under the direction of Dr. M. A. Tinker. The author is grateful for many helpful suggestions.

of mental concentration, insomnia, decreased appetite, headaches and nervousness.

The drug has been used with success in the treatment of narcolepsy by Prinzmetal and Bloomberg (**14**) and by Ulrich, Trapp and Vidgoff (**20**). Schilder (**16**) and Nathenson (**11**) report successful symptomological treatment of neurasthenia and nervous exhaustion with the drug. Wilbur, McLean and Allen (**21**) found it had favorable results with chronic exhaustion, depression and psychoneurosis. Guttmann and Sargent (**6**) conclude that it helped mildly depressed patients, patients in anxiety states and schizophrenics lacking initiative. Some schizophrenics were made more depressed. Reifenstein and Davidoff (**15**) used the drug successfully in the treatment of alcoholic psychotics and they suggest that the results are even more favorable in the case of non-psychotic alcoholics.

Bradley (**5**) gave doses of two m.g. to thirty children, five to fourteen years old, who were in a hospital for behavior disorders. He reports that they became more interested in school, did better work and were emotionally subdued with no loss of interest. He offers the explanation that the apparent inhibition of undesirable behavior was due to the drug stimulating higher cortical centers whose function it is to inhibit behavior.

From these clinical and experimental reports there is some indication that benzedrine sulphate is a stimulant which may increase speed and amount of activity, facilitate higher mental processes and have a direct influence upon affective behavior. There is little indication of any after-effect. The reaction to the drug varies both with the individual and the dosage.

Further investigation of benzedrine in relation to mental and motor tasks is desirable. In this experiment it was decided to determine the influences of fifteen m.g. of benzedrine on blocking during color naming.

Occurrence of blocking in this and other simple responses has been studied by Bills (**1**, **2**, **3**, **4**) who reports that the following principles apply to the phenomena: (*a*) Blocks occur rhythmically, taking the time of two to six responses; (*b*) Practice reduces frequency and size of blocks; (*c*) Fatigue in-

creases frequency and size of blocks; (*d*) Rapid responders have fewer and shorter blocks; (*e*) Errors occur in conjunction with blocks; (*f*) Anoxemia results in more and larger blocks.

Bills suggests that blocking is related to the neurological phenomenon of the cumulative refractory phase and Telford (**19**) has discovered characteristics in associative responses that show much evidence of obeying the laws of the refractory phases. As it seems quite evident that blocking is dependent to a great extent upon the physiological state of the organism, we may expect it to be affected by a changed physiological state. The present experiment was set up to test this hypothesis.

The purposes of the experiment may be more specifically stated: (*a*) To determine the effects of benzedrine sulphate on blocking in regard to number of blocks, length of blocks, location of blocks and effect of fatigue on blocks in a simple response situation (color naming); (*b*) To determine the effects of benzedrine sulphate on the speed of response and the constancy of this speed within experimental trials in the same task. The plan is to account for as much of the variation as possible in the above phenomena and reduce the experimental error to a minimum.

The subjects were eight male and seven female senior and graduate students in psychology. All were in normal health and had normal color vision. Ages of the subjects ranged from seventeen to thirty-two.

The response studied consisted of simple naming of colors from the Woodworth-Wells Color Card. On this card there are 100 colored squares, yellow, blue, black, red and green. Each color was 1 c.m. square. They were arranged in ten rows and ten columns. The card was cut to a perfect square so when the 100 colors were once named, it could be turned on its side and thus continuous responses obtained without repeating sequences. A telegraph key which the subject pressed as he made each response was connected to a Gaertner Chronograph. Responses, time in seconds, and errors were recorded on a moving wax tape. The experimenter stood behind a small screen before the subject and turned the card at every 100 responses and recorded the errors. The recording apparatus was in another room where the noise would not disturb the subject. The experiment was conducted in the afternoon between 1:30 and 4:00.

There were four trials for each subject. Each trial was ten minutes long. The first was a practice trial. The second trial was also alike for all subjects. In the third trial the subjects were divided into three groups—chosen at random. Five were given fifteen m.g. of benzedrine sulphate, five were given placebos (control dosage of sugar tablets), and five were given nothing. The first three trials were given on three successive days. The fourth trial which was alike for all groups was given from four days to two weeks after the third trial. Instructions for each trial and for taking the drug

were printed and alike for all subjects. The drug was distributed to the subjects the day before the experimental or third trial with directions to eat a light lunch the next day and to take the drug two hours before the experimental period.

After the third trial the first two groups were given questionnaires concerning the effect of the drug. The benzedrine group reported that they talked more than usual, accomplished more and concentrated easier. None of the placebo group reported these effects were present. However, all but one of the placebo group later reported they were certain they had been given the drug.

In reading the records, a block was counted as any pause between two successive responses that was equal to or greater than twice the average time between successive responses for each series of 100. Thus the definition of a block was dependent upon the individual's speed during a limited time. Responses were counted and measured to the nearest fifth of a second. Tables I and II present the proportion of responses and the time per response for each group and each trial.

The results were analyzed using Fisher's method of analysis of variance and Snedocor's F test (**18**). These results were checked by 't' tests and critical ratios, using the SE of proportions.

TABLE I

PROPORTION OF RESPONSES WHICH WERE BLOCKS FOR EACH INDIVIDUAL IN EACH GROUP AND EACH TRIAL

Individual	First Control Trial	Experimental Trial	Second Control Trial
	Benzedrine Group		
1	.0549	.0602	.0393
2	.0611	.0683	.0712
3	.0313	.0330	.0333
4	.0603	.0479	.0632
5	.0442	.0655	.0722
	Placebo Group		
1	.0919	.0966	.1038
2	.0488	.0583	.0522
3	.0451	.0450	.0570
4	.0294	.0398	.0522
5	.0596	.0607	.0589
	Control Group		
1	.0490	.0587	.0575
2	.0381	.0397	.0535
3	.0568	.0615	.0731
4	.0750	.0928	.0709
5	.0588	.0586	.0542

TABLE II

AVERAGE TIME IN SECONDS FOR RESPONSES OF EACH INDIVIDUAL IN EACH GROUP AND EACH TRIAL

Individual	First Control Trial	Experimental Trial	Second Control Trial
Benzedrine Group			
1	.75	.62	.62
2	.81	.64	.70
3	.57	.50	.50
4	.70	.62	.66
5	.73	.64	.66
Placebo Group			
1	.83	.69	.77
2	.69	.82	.67
3	.63	.58	.54
4	.60	.58	.54
5	.53	.52	.51
Control Group			
1	.59	.56	.54
2	.65	.66	.61
3	.46	.43	.44
4	.79	.69	.78
5	.67	.60	.57

With the experiment designed in this way there were several sources of variation which could be directly accounted for: (1) the variance between the three groups for each of the three trials; (2) the variance between trials for each of the three groups; (3) the variance between individuals, and (4) the variance between successive responses of each individual in each trial. To determine the effect of the drug a study was made of (*a*) the variance between the experimental trial of the benzedrine group and the first and second control trials of the same group, and (*b*) the variance between the experimental trial of the benzedrine group and the experimental trials of the two control groups. If the benzedrine has an effect, and other things are equal, the responses of the drugged group in the experimental trial should differ from their own responses in the first and second control trials, and the responses of the drugged group should likewise differ from the responses of the two control groups in the experimental trial.

The data were first studied to see if the benzedrine had any effect on the proportion of responses that were blocks. The variance between the three groups in the first control trial and the second control trials in which conditions were alike for all groups was analyzed. The variance was not significant and we thus have evidence that the groups are essentially the same in regard to the proportion of blocks before any differential condition was introduced. The variance between groups in the experimental trial in which there was variation in conditions was also found to be non-significant. We can say, therefore, that the differential treatment (benzedrine and placebo) had no effect on the proportion of responses which are blocks. To substantiate this the variance between the two control trials and the experimental trial for each of the three groups was studied and found to be non-significant. Thus no significant practice effect was present. This may be due to the preliminary ten minute practice period. Further analysis revealed that the greatest source of variation throughout the experiment was the individual differences.

The lengths of the blocks were analyzed in the same way as the frequency of blocks. No significant differentiation was found to exist for the length of blocks.

The proportion of blocks in the first 300 responses and the proportion in the last 300 were selected for each individual in each trial. No significant variation between these six groups of 100 was found either for the placebo or for the control groups in any of the trials. Thus no significant fatigue effect was present. The probability of significance of variance between groups of 100 reached the 5 percent level for the benzedrine group in the first control trial, thus showing a possible fatigue effect. In the experimental trial and the second control trial this variance was not significant. We can conclude, therefore, that there was no significant fatigue effect influencing frequency of blocks in any trial.

The lengths of blocks were analyzed in the same way as the frequency of blocks to discover possible effects of fatigue. No consistent trends were found and the large experimental errors reveal the unreliability of this part of the experiment.

With regard to the average length of time per response, no significant variation between groups was found in either of the two control trials. The variance was not significant between groups in the experimental trials. However, when we compare trials within groups, we find all three groups show an increase in the number of responses or speed, in the experimental trial, but only the increase shown by the benzedrine group is significant. Thus comparison of the experimental trials with the control trials suggests that the benzedrine does result in an increased speed of response.

The speed of responses in the first three hundred and last three hundred responses in each trial was also analyzed for evidence of any fatigue effect. No statistically significant fatigue effect was discovered in either the control or experimental trials.

The analysis of variance method employed in this study reveals the sources of variation. The variance between groups was measured. Then the variance within groups was broken down into variance between trials and variance between individuals. The variance within individuals was broken down into variance within trials and experimental error. This showed a satisfactorily small experimental error in all cases but that part of the experiment dealing with length of blocks.

The data warrant the following conclusions: (1) Fifteen m.g. of benzedrine sulphate has no effect on (*a*) number of blocks occuring in color naming, (*b*) length of blocks, (*c*) location of blocks within trials, (*d*) influence of fatigue on length of blocks, and (*e*) influence of fatigue on speed of response; (2) Fifteen m.g. benzedrine sulphate slightly increases speed of response in color naming; (3) Most of the variation in blocking and speed of response is due to individual differences.

(Manuscript received April 10, 1940)

References

1. Bills, A G., Blocking: A new principle of mental fatigue, *Amer. J. Psychol.*, 1931, **43**, 230–245.
2. Bills, A. G., Blocking in mental fatigue and anoxemia compared, *J. exp. Psychol.*, 1937, **20**, 437–452.
3. Bills, A. G., Fatigue, oscillation and blocks, *J. exp. Psychol.*, 1935, **18**, 562–573.

4. Bills, A. G., Some causal factors in mental blocking, *J. exp. Psychol.*, 1935, **18**, 172–185.
5. Bradley, C., The behavior of children receiving benzedrine, *Amer. J. Psychiat.*, 1937, **94**, 577–585.
6. Guttmann, E., and Sargant, W., Observations on benzedrine, *Brit. Med. J.*, 1937, **1**, 1013–1015.
7. Gwynn and Yates, Benzedrine sulphate as a mental stimulant, *Med. Ann. D. C.* 1937, **6**, 356.
8. McNamara, W. J., and Miller, R. E., Effects of benzedrine sulphate on mental work, *Psychol. Rec.*, 1937, **7**, 78–84.
9. Molitch, M., and Eccles, A. K., The effect of benzedrine sulphate on the intelligence scores of children, *Amer. J. Psychiat.*, 1937, **94**, 587 and 590.
10. Molitch, M., and Sullivan, J. P., The effect of benzedrine sulphate on children taking the new Stanford Achievement Tests, *Amer. J. Orthopsychiat.*, 1937, **7**, 519–522.
11. Nathanson, M. H., The central action of beta-aminopropyl benzene, *J. A. M. A.*, 1937, **108**, 328–531.
12. O'Connor, D. M., Letter in *Brit. Med. J.*, 1937, **1**, 86.
13. Peoples and Guttmann, Hypertension produced with benzedrine: Its psychological accompaniments, *Lancet*, 1936, **20**, 1107–1109.
14. Prinzmetal, M., and Bloomberg, W., The use of benzedrine for the treatment of narcolepsy, *J. A. M. A.*, 1935, **105**, 2051–2054.
15. Reifenstein, E. C., and Davidoff, E., The treatment of alcoholic psychoses with benzedrine sulphate, *J. A. M. A.*, 1938, **110**, 1811–1812.
16. Schilder, P., The psychological effect of benzedrine sulphate, *J. Nerv. & Ment. Dis.*, 1938, **87**, 584–587.
17. Simpson, S. L., Letter, *Brit. Med. J.*, 1937, **1**, 93.
18. Snedecor, G. W., *Statistical Methods*, Collegiate Press, Inc., Ames Iowa, 1937.
19. Telford, C. W., The refractory phase of voluntary and associative response, *J. exp. Psychol.*, 1931, **14**, 1–36.
20. Ulrich H., Trapp, C. E., and Vidgoff, B., The treatment of narcolepsy with benzedrine sulphate, *Ann. Internat. Med.*, 1939, **9**, 1213–1221.
21. Wilbur, D. L., McLean, A. K., and Allen E. V., Clinical observations on the effect of benzedrine sulphate, *J. A. M. A.*, 1937, **109**, 549–554.

A HEMI-PERISPHERE FOR VISUAL MEASUREMENTS

BY ROBERT W. BURNHAM

Department of Psychology,
Rutgers University, New Brunswick, N. J.

In the measurement of the limits of the retinal color fields, visual blind spot, and other peripheral visual functions, apparatus in the past has been limited usually to some form of semi-circular arc arrangement. Devices of that type have reached virtually the limit of mechanical perfection still without providing an efficient means for obtaining a homogeneous background devoid of distractions, or without well controlled illumination of the 'field.'

A device has been designed and a model constructed which has all the advantages of older types of perimeters, and which has practically eliminated the disadvantages. This particular device was constructed to measure color zones. A virtual 'Ganzfeld' is available to the observer, there being a homogeneous ground obtained by the use of a thin hemispherical aluminum shell,[1] 20 inches in diameter, mounted at its center on a bearing by means of a steel rod projecting to the outside of the shell from the center axial point of the shell. This steel rod screws into the shell at this point and serves inside as a fixation point (see Fig. 1).

Upon the open edge of the shell at its locus of maximum diameter a cylindrical front portion was constructed by bending a sheet of thin aluminum, 3 inches wide, into a cylinder having an inside diameter of 20 inches so that it could be fitted over the edge of the shell. A heavy circular copper band of rectangular cross section $\frac{1}{8}$ inch $\times$ $\frac{1}{2}$ inch was bolted over the cylinder-hemisphere joint. The rear edge of this band of copper serves as a 'track' upon which the cylinder-hemisphere unit can rotate. The band also gives strength to the unit.

[1] These shells can be obtained relatively cheaply in an unfinished form, by writing to the author of this article.

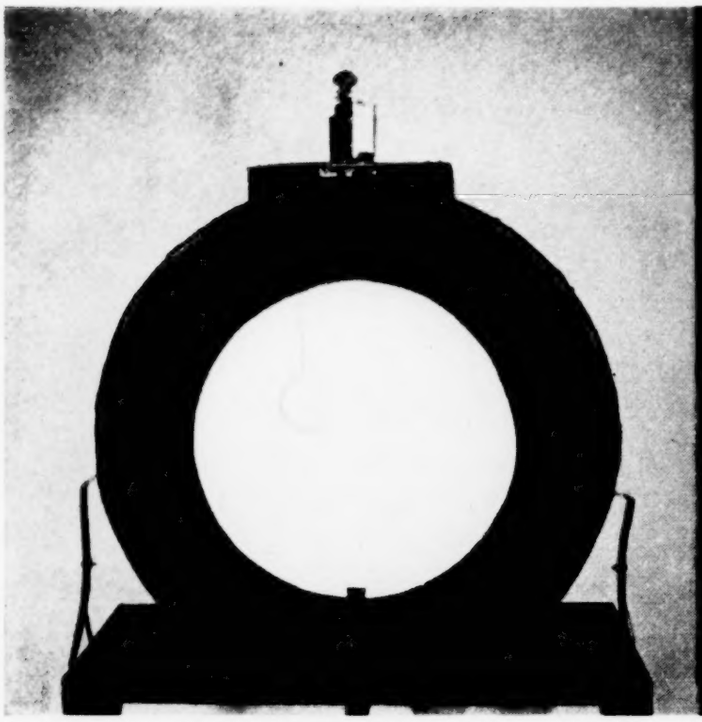


FIG. 1

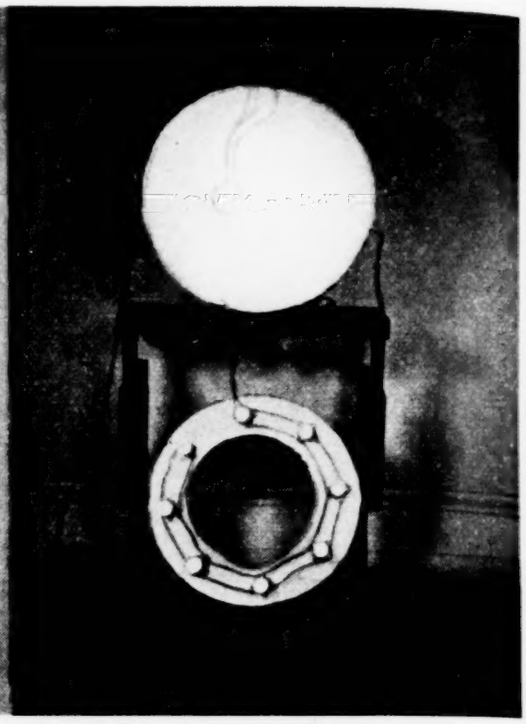


FIG. 2

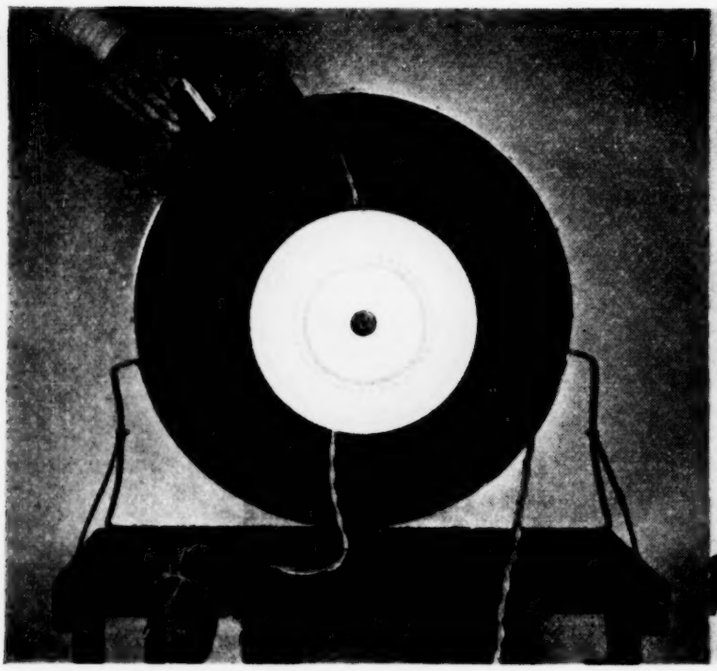


FIG. 3

FIG. 4

A thin circular aluminum facing, 20 inches in diameter, was cut out and bolted to the cylindrical front of the cylinder-hemisphere unit. In the center of this circular facing was cut a round hole, 14 inches in diameter, inside of which was

fastened another cylinder similar to that outside. (This is most plainly visible in Fig. 4.) The whole front-piece consisting of the two cylinders and facing looks very much like a large doughnut-cutter when removed from the hemisphere. The facing and inner cylinder, unbolted from the outer cylinder, are shown at the bottom of Fig. 2. Inside this cylindrical front-piece eight lights were arranged (see Fig. 2) to provide controlled illumination of the field. Small bulbs of high wattage radiating little heat are necessary in this limited enclosed space.

On top of the shell was mounted a protractor on a heavy brass semi-circle, and at the center of the protractor hangs a hollow copper tube, $\frac{1}{2}$ inch inside diameter, which was bent to conform to the shape of the shell. At the lower end of this tube, as shown in Fig. 2, a target was fastened. This target consists of a circular piece of copper which has at its center a 2 mm circular aperture, serving as a stimulus patch. On the back of the target was placed a track to hold vertically a series of four small Wratten filters mounted in a row in glass. A small electric bulb (having a dry-cell source) placed behind the filter track provided illumination for the filters. The proper filter could be moved into position for observation by pulling upward on a knob at the upper end of the copper tube (outside the shell) which was fastened to an automobile choke-wire that passed down through the copper tube and was fastened at its lower end to the row of mounted filters. Wiring for the electric bulb was accomplished by using the copper tube as one conductor and the choke wire for the other. This was possible since the copper tube was lined with rubber tubing which serves as insulation. A sturdy steel bearing through which the copper tube runs was mounted at the center of the protractor. The heavy bearing serves to keep the copper tube and target steady and rotating on a single axis. Attached to the upper end of the copper tube (which projected above the bearing) is a pointer-arm to indicate degrees on the protractor. This arm also is used to turn the target.

The shell rests at the front upon a semicircular track of copper fastened to the stand upon which the whole apparatus

was placed and having the same cross-sectional dimensions as the track on the cylinder-hemisphere unit. The track on the cylinder-hemisphere unit was placed in front of and next to the track upon which the shell rested. The shell was held upright in this position (where it can be rotated) by a collar placed on the steel tube located at the rear center of the shell. This steel tube passes through a bearing mounted by means of a suitable support upon the stand. The collar is on the back side of the bearing and so holds the shell firmly between the bearing (on the back) and the track (on the front). On the rearmost end of the steel tube which passes through the bearing was placed a circular aluminum disc (see Fig. 3) calibrated in degrees and of the type used on color-mixing motors. This disc is faced by a fixed pointer to indicate degrees of turning of the shell. It is possible to rotate the shell to any desired position, bringing the protractor scale and attached copper tube with target and stimulus patch to any meridian in which it is desired to make measurement. The whole apparatus is mounted upon a stand having legs which can be adjusted to any height depending upon the subject. A head rest is mounted on the stand. The steel base of the head rest curves around and into the shell so that the subject's eye for observation is directly in line with the center of the turning copper tube with attached stimulus. The inside of the shell, the copper tube, the target, and the head rest were sprayed with white paint giving a pronounced effect of depth with a minimum of distracting elements. Shadows are negligible and the target blends with the ground almost completely.

It can readily be seen from the illustrations that many refinements are possible on this apparatus. The present one serves merely as a model. It eliminates customary sources of distraction, gives controlled background illumination as well as a homogeneous field almost lacking in apparent surface, and it may be varied to measure all perimetric visual functions. The apparatus is also relatively inexpensive to construct; the materials for this particular model cost less than $30.

(Manuscript received April 9, 1940)

A SIMPLE, INEXPENSIVE, AND PORTABLE APPARATUS FOR DEMONSTRATING THE 'PHANTOM' SOUND

BY HILDING B. CARLSON

The University of Illinois

If two similar sounding objects are located differently and vibrated simultaneously, a single sound—known as the phantom sound—is heard and is localized as being at some position between the two objects. Within certain limits, the apparent location of the resulting sound can be varied by a number of techniques, one of which is to alter the relative intensities of the vibrations of the two objects. This is the method that was used by Matsumota (**2**) and others in studies on the directional localization of sound, and is the method employed in the apparatus to be described.

In order to demonstrate the phantom sound to students in a course in perception, the author devised a simple, inexpensive and portable apparatus. Since the entire mechanism can be enclosed in a box less than 12″ × 6″ × 5″ in size, it is small enough to be transported easily to and from a class room, so that demonstrations can readily be given in class rooms in schools that do not possess separate rooms for experimental and demonstrational purposes. The essential parts include a pair of ear phones (1500 ohms), a potentiometer R_1 (200,000 ohms) with uniform resistance change, another potentiometer R_2 (2,000 ohms), an ordinary buzzer, a flashlight battery, a switch, and wire. Most of the materials used in this apparatus were obtained from a discarded radio set. In order to diminish the sound at the buzzer, the entire apparatus was enclosed in a box made of sound absorbing boards. The hook-up is given in Diagram 1.

In operation, the ear phones are snugly adjusted to the head, the switch is turned on, rheostat R_2 is operated so as to get an intensity of sound neither too loud for comfort nor too

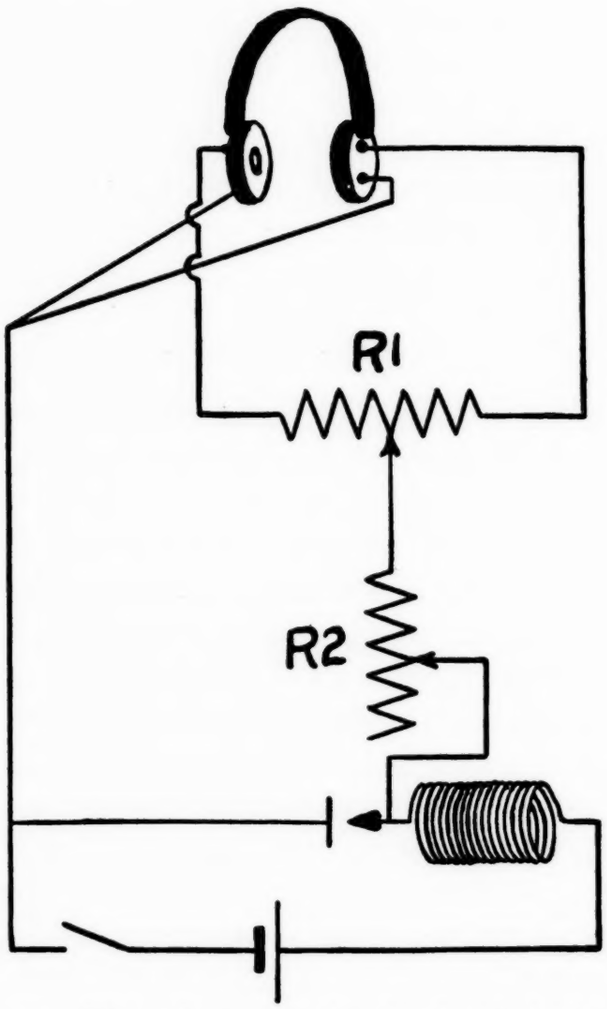


DIAGRAM 1. Wiring diagram for changing the relative intensities of the sound at the two ears.

weak to prevent getting a marked phantom sound effect, rheostat R_1 is then rotated slowly, first in one direction to the extreme degree, then in the other direction to its extreme position, etc. The locus of the resulting sound will be perceived as moving from one ear to the other, back again, etc. Some subjects report that they localize the sound as within their heads, others as above their heads, and some as behind their heads. Further discussion of this phenomenon is given in (1).

(Manuscript received April 19, 1940)

REFERENCES

1. CARR, H. A., *An Introduction to Space Perception*, Longmans, Green, N. Y., 1935.
2. MATSUMOTA, Researches on acoustic space, *Yale Studies*, 1897, 5, 1–75.